

## DIPLOMATURA D'ESTADÍSTICA

---

### **Modelat de la Gravetat d'Accidents Industrials amb Substàncies Perilloses**

*Alumnes:* Gina Albero Abril  
M<sup>a</sup> Ángeles Martínez Turégano

*Director:* Josep Ginebra Molins  
*Departament:* Estadística i Investigació Operativa  
*Data d'entrega:* 4 de juliol de 2.000

---

UNIVERSITAT POLITÈCNICA DE CATALUNYA  
Biblioteca



1400708221

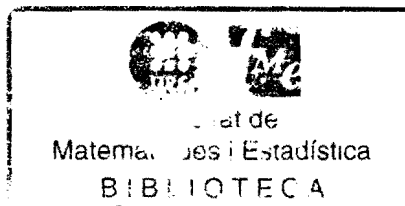


Facultat de Matemàtiques  
i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Qualificació:

Excel·lent 9.



PRESIDENT

Prof. T. Aluja

SECRETARI

Prof. X. Heredia

VOCAL

Prof. A. Avilés

Barcelona, 19 de juliol de 2000

*Agraïm al nostre director Josep Ginebra la direcció d'aquest projecte, el suport i els consells donats. Sobretot perquè ens ha dedicat part del seu valuós temps i ens ha ajudat a aplicar els coneixements rebuts en la Diplomatura d'Estadística. A l'enginyer químic Sergio Carol per haver confiat en nosaltres per desenvolupar aquest projecte i l'amabilitat amb la qual ens ha resolt els dubtes sobre el tema estudiat. A les nostres famílies per la seva comprensió i suport prestat durant el temps de realització del projecte. A les persones que treballen en el laboratori de càlcul de la FME per resoldre tots els problemes informàtics sorgits. I en general, a totes les persones que ens han recolzat al llarg de la realització d'aquest projecte.*

# ÍNDEX

|  |           |
|--|-----------|
| <b>CAPÍTOL 1: INTRODUCCIÓ .....</b>  | <b>4</b>  |
| <b>CAPÍTOL 2: DESCRIPCIÓ DE LA BASE DE DADES MHIDAS.....</b>   | <b>7</b>  |
| 2.1. INTRODUCCIÓ.....  | 7         |
| 2.2. DESCRIPCIÓ DE LES VARIABLES RESPOSTA .....  | 11        |
| 2.2.1. <i>Variables que descriuen el dany a les persones</i> .....   | 11        |
| 2.2.2. <i>Variable DAMAGE</i> .....  | 12        |
| 2.3. DESCRIPCIÓ DE LES VARIABLES EXPLICATIVES .....  | 13        |
| 2.3.1. <i>Variables que indiquen la data de l'accident</i> .....   | 13        |
| 2.3.2. <i>Variables geogràfiques</i> .....   | 14        |
| 2.3.3. <i>Variables que indiquen les substàncies involucrades</i> .....  | 15        |
| 2.3.4. <i>Variables que descriuen el tipus d'accident</i> .....  | 18        |
| 2.3.5. <i>Variables que expliquen les causes de l'accident</i> .....   | 20        |
| 2.3.6. <i>Variable que mesura la quantitat de substància involucrada</i> .....   | 26        |
| 2.3.7. <i>Variables que expliquen l'accident</i> .....   | 26        |
| 2.3.8. <i>Variables complementàries</i> .....  | 27        |
| 2.4. RESUM DEL MÉS IMPORTANT .....   | 28        |
| <b>CAPÍTOL 3: ADAPTACIÓ DE LA BASE DE DADES MHIDAS.....</b>  | <b>31</b> |
| 3.1. INTRODUCCIÓ.....  | 31        |
| 3.2. DEPURACIÓ DE LES VARIABLES RESPOSTA .....   | 32        |
| 3.2.1. <i>Depuració de les variables nombre de morts, nombre de ferits i nombre d'evacuats</i> .....                           | 32        |
| 3.2.2. <i>Depuració de la variable danys econòmics</i> .....   | 36        |
| 3.3. DEPURACIÓ DE LES VARIABLES EXPLICATIVES.....  | 38        |
| 3.3.1. <i>Depuració de la data de l'accident</i> .....   | 38        |
| 3.3.2. <i>Depuració de les variables geogràfiques</i> .....  | 39        |
| 3.3.3. <i>Depuració de les variables que indiquen les característiques de les substàncies involucrades en l'accident</i> ..... | 42        |
| 3.3.4. <i>Depuració de les variables que descriuen el tipus d'accident</i> .....   | 45        |
| 3.3.5. <i>Depuració de les variables que expliquen les causes de l'accident</i> .....  | 47        |
| 3.3.6. <i>Depuració de la variable que mesura la quantitat de substància involucrada en l'accident</i> .....                   | 50        |
| 3.3.7. <i>Depuració de les variables que expliquen l'accident</i> .....  | 51        |
| 3.3.8. <i>Depuració de les variables complementàries</i> .....   | 51        |
| 3.4. RESUM DEL MÉS IMPORTANT .....   | 58        |
| <b>CAPÍTOL 4: ANÀLISI EXPLORATÒRIA DE DADES .....</b>  | <b>62</b> |
| 4.1. INTRODUCCIÓ.....  | 62        |
| 4.2.1. <i>Variable Killed i transformació</i> .....  | 64        |
| 4.2.2. <i>Variable Injured i transformació</i> .....   | 66        |
| 4.2.3. <i>Variable Evacuated i transformació</i> .....   | 68        |
| 4.2.4. <i>Variable danys econòmics i transformació</i> .....   | 70        |
| 4.3. ANÀLISI UNIVARIANT DE LES VARIABLES EXPLICATIVES .....  | 72        |



|   |            |
|---|------------|
| 4.3.1. Variables relacionades amb la data de l'accident .....                           | 72         |
| 4.3.2. Variables geogràfiques .....   | 74         |
| 4.3.3. Variables que indiquen les substàncies involucrades .....                        | 74         |
| 4.3.4. Variables que descriuen el tipus d'accident .....                                | 77         |
| 4.3.5. Variables que expliquen les causes de l'accident .....                           | 78         |
| 4.3.6. Variable que mesura la quantitat de substància involucrada i transformació ..... | 79         |
| 4.4. ANÀLISI BIVARIANT PER KILLED .....   | 81         |
| 4.4.1. Relació de Killed amb la data de l'accident .....                                | 81         |
| 4.4.2. Relació de Killed amb la situació geogràfica .....                               | 83         |
| 4.4.3. Relació de Killed amb les substàncies involucrades .....                         | 85         |
| 4.4.4. Relació de Killed amb el tipus d'accident .....                                  | 89         |
| 4.4.5. Relació de Killed amb el tipus d'activitat .....                                 | 91         |
| 4.4.6. Relació de Killed amb la quantitat de substància involucrada .....               | 94         |
| 4.5. ANÀLISI BIVARIANT PER INJURED .....  | 95         |
| 4.5.1. Relació de Injured amb la data de l'accident .....                               | 95         |
| 4.5.2. Relació de Injured amb la situació geogràfica .....                              | 97         |
| 4.5.3. Relació de Injured amb les substàncies involucrades .....                        | 98         |
| 4.5.4. Relació de Injured amb el tipus d'accident .....                                 | 102        |
| 4.5.5. Relació de Injured amb el tipus d'activitat .....                                | 104        |
| 4.5.6. Relació de Injured amb la quantitat de substància involucrada .....              | 107        |
| 4.6. ANÀLISI BIVARIANT PER DAMAGE .....   | 108        |
| 4.6.1. Relació de Damage amb la data de l'accident .....                                | 108        |
| 4.6.2. Relació de Damage amb la situació geogràfica .....                               | 111        |
| 4.6.3. Relació de Damage amb les substàncies involucrades .....                         | 112        |
| 4.6.4. Relació de Damage amb el tipus d'accident .....                                  | 115        |
| 4.6.5. Relació de Damage amb el tipus d'activitat .....                                 | 116        |
| 4.6.6. Relació de Damage amb la quantitat de substància involucrada .....               | 119        |
| 4.7. CONCLUSIONS .....  | 120        |
| <b>CAPÍTOL 5: MODELS LINEALS .....</b>  | <b>123</b> |
| 5.1. INTRODUCCIÓ .....  | 123        |
| 5.2. TAULA DE LES VARIABLES RESPOSTA .....  | 126        |
| 5.3. TAULES DE LES VARIABLES EXPLICATIVES .....   | 127        |
| 5.4. ESTRATÈGIA A SEGUIR EN EL MODELAT .....  | 130        |
| <b>CAPÍTOL 6: MODEL LINEAL PER A KILLED .....</b>                                       | <b>133</b> |
| 6.1. INTRODUCCIÓ .....  | 133        |
| 6.2. MODELS PELS ACCIDENTS AMB MORTS .....  | 134        |
| 6.2.1. Model amb la variable CAUSA GENERAL .....  | 135        |
| 6.2.2. Model sense la variable CAUSA GENERAL .....                                      | 139        |
| 6.3. INTERPRETACIÓ DELS MODELS PELS ACCIDENTS AMB MORTS .....                           | 143        |
| 6.3.1. Model amb la variable CAUSA GENERAL .....  | 143        |
| 6.3.2. Model sense la variable CAUSA GENERAL .....                                      | 144        |
| 6.4. MODELS PER TOTS ELS ACCIDENTS (INCLOSOS ELS SENSE MORTS) .....                     | 146        |
| 6.4.1. Model amb la variable CAUSA GENERAL .....  | 147        |
| 6.4.2. Model sense la variable CAUSA GENERAL .....                                      | 150        |
| 6.5. INTERPRETACIÓ DELS MODELS PER TOTS ELS ACCIDENTS (INCLOSOS ELS SENSE MORTS) .....  | 153        |
| 6.5.1. Model amb la variable CAUSA GENERAL .....  | 153        |
| 6.5.2. Model sense la variable CAUSA GENERAL .....                                      | 155        |
| 6.6. CONCLUSIONS .....  | 156        |

**CAPÍTOL 7: MODEL LINEAL PER A INJURED..... 157**

|  |     |
|--|-----|
| 7.1. INTRODUCCIÓ.....  | 157 |
| 7.2. MODELS PELS ACCIDENTS AMB FERITS.....   | 158 |
| 7.2.1. <i>Model amb la variable CAUSA GENERAL</i> .....                                | 159 |
| 7.2.2. <i>Model sense la variable CAUSA GENERAL</i> .....                              | 162 |
| 7.3. INTERPRETACIÓ DELS MODELS PELS ACCIDENTS AMB FERITS.....                          | 166 |
| 7.3.1. <i>Model amb la variable CAUSA GENERAL</i> .....                                | 166 |
| 7.3.2. <i>Model sense la variable CAUSA GENERAL</i> .....                              | 167 |
| 7.4. MODELS PER TOTS ELS ACCIDENTS (INCLOSOS ELS SENSE FERITS).....                    | 169 |
| 7.4.1. <i>Model amb la variable CAUSA GENERAL</i> .....                                | 170 |
| 7.4.2. <i>Model sense la variable CAUSA GENERAL</i> .....                              | 173 |
| 7.5. INTERPRETACIÓ DELS MODELS PER TOTS ELS ACCIDENTS (INCLOSOS ELS SENSE FERITS)..... | 175 |
| 7.5.1. <i>Model amb la variable CAUSA GENERAL</i> .....                                | 175 |
| 7.5.2. <i>Model sense la variable CAUSA GENERAL</i> .....                              | 177 |
| 7.6. CONCLUSIONS.....  | 178 |

**CAPÍTOL 8: MODEL LINEAL PER A DAMAGE ..... 179**

|   |     |
|---|-----|
| 8.1. INTRODUCCIÓ.....   | 179 |
| 8.2. MODEL AMB LA VARIABLE CAUSA GENERAL.....                     | 181 |
| 8.3. INTERPRETACIÓ DEL MODEL AMB LA VARIABLE CAUSA GENERAL.....   | 183 |
| 8.4. MODEL SENSE LA VARIABLE CAUSA GENERAL.....                   | 184 |
| 8.5. INTERPRETACIÓ DEL MODEL SENSE LA VARIABLE CAUSA GENERAL..... | 186 |
| 8.6. CONCLUSIONS.....   | 187 |

**CAPÍTOL 9: CONCLUSIONS ..... 188**

**VALORACIÓ TEMPORAL I ECONÒMICA..... 198**

**ANNEX A: MOSTRA DE LA BASE DE DADES MHIDAS ..... 199**

**ANNEX B: MOSTRA DE LA BASE DE DADES MHIDAS ADAPTADA ..... 203**

**ANNEX C: MOSTRA DE LA BASE DE DADES UTILITZADA PER LA  
MODELITZACIÓ..... 207**

**BIBLIOGRAFIA ..... 214**

# CAPÍTOL 1: INTRODUCCIÓ

En aquest projecte de final de carrera s'estudien els accidents industrials on han intervingut substàncies químiques perilloses. Aquests accidents estan recollits en la base de dades anglesa que és comercialitzable i porta com a nom MHIDAS (Major Hazard Incident Data Service). La versió que es disposa d'aquesta base de dades és la de juliol de 1.999 i conté 10.102 substàncies químiques perilloses que han participat en 8.790 accidents industrials.

El treball de recopilació es va començar a realitzar a principis dels anys 80, encara que es registren accidents anteriors a l'any 1.900. Per l'estudi es disposa de 8.790 accidents industrials amb substàncies perilloses, que es descriuen mitjançant variables contínues i categòriques, i on cada registre es correspon a un únic accident.

Inicialment, la base de dades va ser proporcionada en format EXCEL. Com que es podia necessitar la base de dades inicial, on cada observació és una substància que ha participat en un accident, posteriorment va ser facilitada una base de dades en format ACCESS, que conté dues taules. Una taula llista cada una de les substàncies que havien participat en un accident, fins a un total de 10.102 registres, i una altra taula llista accidents, on hi ha participat substàncies químiques perilloses. En un mateix accident hi poden haver intervingut més d'una substància química. Aquesta última taula conté 8.790 registres, que corresponen a 8.790 accidents.

Per fer els anàlisis pròpiament estadístics s'utilitzarà el paquet estadístic MINITAB. Això vol dir que en aquest projecte s'utilitzaran tres programes informàtics, aprofitant les propietats que tenen cadascun d'ells, per desenvolupar cada una de les tasques determinades.

Aquest estudi ha sorgit a partir de la Tesi Doctoral que porta per títol provisional “Desarrollo de un índice de riesgo tecnológico a partir de la aplicación del análisis histórico de accidentes con sustancias peligrosas” i que està realitzant l'enginyer químic Sergio Carol Llopart sota la direcció del professor Dr. Juan Antonio Vilchez del Departament d'Enginyeria Química de la Universitat Politècnica de Catalunya, de Barcelona. Aquesta tesi s'està realitzant al CERTEC, Centre d'Estudis del Risc Tecnològic, que està situat al pavelló G de l'Escola Tècnica Superior d'Enginyeria Industrial de Barcelona.

El propòsit inicial és conèixer i familiaritzar-se amb la base de dades que es disposa, per tal d'aplicar les tècniques estadístiques més adequades a la problemàtica en qüestió. Cal destacar que la base de dades és molt gran perquè té molts registres i una gran quantitat de variables, la majoria qualitatives i amb moltes categories, moltes dades faltants i molts errors de codificació i sintaxi que s'hauran de corregir i ordenar. Això implica dedicar molt de temps a realitzar una depuració profunda i delicada de la base de dades, i a simplificar el gran nombre de categories que tenen les variables qualitatives, a base d'agrupar-les en categories semblants.

Tot això es farà de la millor manera possible per tal de no perdre informació, no esbiaixar els resultats i poder aprofitar al màxim la informació de la que es disposa per tal d'obtenir una base de dades que sigui el més fiable possible. Aquesta serà la part més difícil i formativa d'aquest projecte perquè obligarà a treballar molt a prop de l'enginyer responsable del projecte que ha sol·licitat la nostra ajuda per realitzar tasques d'estadística.

L'objectiu final del projecte és modelar la gravetat dels accidents a partir de l'anàlisi històric d'accidents industrials amb substàncies químiques perilloses. Com a variables resposta que reflexen la gravetat de l'accident es farà servir el nombre de morts, el nombre de ferits, el nombre de persones evacuades i l'estimació econòmica del dany material.

Abans de construir els models de regressió per aquestes 4 variables, es farà una anàlisi exploratòria de les dades recollides amb l'objectiu d'evidenciar les relacions existents entre els factors que intervenen en els accidents industrials amb substàncies químiques perilloses i les seves conseqüències.

S'haurà de dedicar una part molt important del temps a analitzar els resultats extrets de la modelització i a analitzar la bondat dels resultats.

Degut a que els accidents industrials amb substàncies perilloses i les variables que els expliquen no es coneix a fons, s'haurà de dedicar també bastant temps parlant amb l'enginyer químic responsable del projecte, per tal d'abordar i solucionar els problemes que es trobaran al llarg de la seva realització.

# **CAPÍTOL 2: DESCRIPCIÓ DE LA BASE DE DADES MHIDAS**

## **2.1. Introducció**

En aquest capítol es descriu l'estructura i contingut de la base de dades que es disposa per l'estudi. Aquesta ha estat obtinguda a partir de la base de dades MHIDAS (Major Hazard Incident Data Service) que va ser creada pel SDR (Safety and Reliability Directorate) que pertany al HSE (Health and Safety Executive). MHIDAS descriu els accidents industrials produïts per substàncies perilloses mitjançant el registre d'aquells factors rellevants en la seva ocurrència (descripció de la instal·lació d'origen, propietats bàsiques de les substàncies i conseqüències). Per la descripció d'un accident MHIDAS utilitza tants registres com substàncies diferents hagin intervingut en el mateix. En canvi, la base de dades que es disposa per l'estudi classifica les substàncies de MHIDAS per accidents.

Les dades es van començar a recollir a principis de 1.980, però s'hi poden trobar accidents anteriors a l'any 1.900 i arriben fins l'any 1.998. Han estat obtingudes a partir de diferents publicacions (informes tècnics, investigacions judicials o premsa especialitzada) sobre accidents industrials que han succeït en qualsevol part del món, tot i que la majoria d'accidents són dels Estats Units (2.795 accidents) i del Regne Unit (2.601 accidents). Una gran part de les dades s'han recollit de la premsa la qual cosa comporta la presència de moltes dades mancants. En la versió de gener de 1.999 la base de dades contenia 9.876 referències de substàncies que havien participat en 8.600 accidents, i en la versió de juliol de 1.999 contenia 10.102 referències de substàncies que havien participat en 8.790 accidents. La base de dades que es disposa per l'estudi conté els accidents d'aquesta última versió.

Actualment, es comercialitza en suport CD-ROM i es distribueix conjuntament amb el programa WINSPIRS, un gestor de bases de dades que les permet filtrar i manipular. També existeix la possibilitat de fer consultes puntuals directament al SDR. WINSPIRS fa cerques de registres mitjançant criteris lògics bastant limitats. Els resultats de cada selecció (Figura 1) només es poden obtenir en format de fitxer de text o impresos en paper. Per tant, no és possible treballar directament sobre la base de dades MHIDAS.

Com s'ha comentat abans, a MHIDAS cada registre correspon a una substància que ha intervingut en un accident i per tant, s'hi pot trobar varis registres que parlin del mateix accident. Cada substància té un codi per accident, si en un accident hi ha participat més d'una substància, aquestes substàncies tenen el mateix codi acompanyat d'una lletra. Per exemple: 7.342 A, 7.342 B i 7.342 C, és a dir, en l'accident 7.342 hi ha participat 3 substàncies. En la base de dades que es disposa per l'estudi, en cada registre hi ha informació sobre un únic accident, i inclou el nombre i les característiques de totes les substàncies que han intervingut. Per tal de poder relacionar les dues bases de dades, MHIDAS (cada observació és una substància) i la que es disposa per l'estudi (cada observació és un accident), en aquesta última s'ha conservat el codi i s'ha eliminat la lletra. Aquesta classificació s'ha fet amb el programa VISUAL BASIC. S'ha considerat que substàncies amb la mateixa data i localització pertanyen al mateix accident. La base de dades de substàncies que es disposa està en el format ACCESS i la d'accidents en format EXCEL.

```
Record 1 of 1047 - OSH-ROM: MHIDAS
AN: 4925
CR: SRD
DA: 07/07/82
LO: ;; YUGOSLAVIA
MN: GASOLINE
MT: LIQUID
MH: FI-fire
MC: 1270
IT: EXPLODE-explosion; TANKFIRE-tank-fire
OG: STORAGE-storage-plant; ASVESSEL-storage-vessels,-atmospheric-pressure
GC: HUMAN-human-factor; EXTERNAL-external-events
SC: PROCEDURES; TEMPRTURE-temperature-extremes
AB: EXPLOSION IN OPEN SPACE ROUND FLOATING ROOF TANK HOLDING GASOLINE. TANK
UNSUITABLE FOR THIS DUTY AND AT SUMMER TEMPERATURES VOLUME IN THE GAS PHASE
WAS GREATER THAN THE TANK COULD HOLD. CLOUD IGNITED BY PROCESS HEATER
CAUSING EXPLOSION AND TANK FIRE.
RA: 1
```

Figura 1. Resultats del programa WINSPIRS per extreure les dades de MHIDAS.

La base de dades MHIDAS està formada per 28 variables de les quals 6 són numèriques, 14 categòriques i 8 de text. Aquestes variables són:

- Les variables que descriuen el dany a les persones, de text però que es convertiran a numèriques, són les següents:

**KILLED:** nombre de persones mortes

**INJURED:** nombre de persones ferides

**EVACUATED:** nombre de persones evacuades

- La variable que indica l'estimació econòmica del dany material produït per l'accident, és de text però es convertirà també a numèrica, és la següent:

**DAMAGE**

- La data de l'accident es coneix a través de les variables numèriques:

**DIA**

**MES**

**AÑO**

- Les variables categòriques geogràfiques són:

**PD:** densitat de població

**LOCATION:** localitat de l'accident

**REGION:** regió de l'accident

**COUNTRY:** estat de l'accident

- Les variables que donen informació sobre les substàncies involucrades són les següents categòriques:

**MN:** nom de la substància

**MT:** estat físic de la substància

**MH:** risc de la substància

**MC:** codi per identificar la substància o la família a la que pertany

i la numèrica:

**NSUS:** nombre de substàncies que han intervingut en l'accident.



- ❑ Les variables que indiquen el tipus d'accident són les categòriques:
  - IT:** tipus d'accident
  - IS:** font d'igniciói la numèrica:
  - IG:** temps que transcorre des de que s'origina l'incident desencadenant fins l'explosió o incendi
  
- ❑ Les variables categòriques que expliquen la causa de l'accident són:
  - OG:** origen de l'accident
  - GC:** causa general
  - SC:** causa específica de l'accident
  
- ❑ La variable de text, però que es convertirà a numèrica, que indica la quantitat de substància involucrada és:
  - QY.**
  
- ❑ Les variables de text que expliquen els accidents són:
  - AB:** fa un resum de l'accident
  - KW:** petit text amb paraules clau
  
- ❑ La variable complementària numèrica és:
  - AN:** codi de l'accidentla variable categòrica:
  - CR:** font d'informaciói la variable de text:
  - RA:** nombre d'articles disponibles per consultar

A continuació s'explicaran més detalladament aquestes variables.

## **2.2. Descripció de les variables resposta**

Les variables resposta d'interès són: **DAMAGE**, estimació econòmica del dany material i **KILLED**, nombre de morts en l'accident, ja que són les que donen una visió de la gravetat de l'accident. A més a més, també es consideren respostes interessants: **INJURED**, nombre de persones ferides i **EVACUATED**, nombre de persones evacuades perquè també donen una idea de la gravetat de l'accident.

### **2.2.1. Variables que descriuen el dany a les persones**

Les variables de text que s'expliquen a continuació, tenen un problema. A més a més de tenir valors concrets, contenen símbols matemàtics de desigualtats, com per exemple:  $> 20$  i  $< 10$ . Això és degut a que no es coneixia el valor exacte i es va fer una estimació a la baixa o a l'alça. Si es volen utilitzar com a variables resposta s'hauran de transformar a numèriques.

- ❑ **KILLED** indica el nombre de morts en l'accident, o una estimació dels mateixos. Conté 1.846 missings (dades mancants), és a dir, es coneix el nombre de morts de 6.944 accidents.
- ❑ **INJURED** indica el nombre de persones ferides en l'accident, o una estimació dels mateixos. Conté 986 missings, és a dir, es coneix el nombre de ferits de 7.804 accidents.
- ❑ **EVACUATED** indica el nombre de persones evacuades en l'accident, o una estimació dels mateixos. Conté 3.293 missings, és a dir, es coneix el nombre d'evacuats de 5.497 accidents.

### 2.2.2. Variable DAMAGE

- **DAMAGE** és una variable de text formada per nombres, però alguns contenen símbols matemàtics, com per exemple:  $< \times 10E6$ . Fa una estimació econòmica del dany material provocat per l'accident en milions de dòlars. Té 7.850 missings, és a dir, que només es coneix l'estimació econòmica del dany material de 940 accidents. Aquesta variable té el mateix problema que les variables explicades anteriorment, ja que conté símbols de desigualtats (" $<$ " o " $>$ ") i això dificultarà la seva transformació a variable numèrica. Si es vol comparar el dany material d'accidents de diferents anys s'haurà de fer una actualització econòmica, ja que no és el mateix un dòlar dels anys 30 que dels 70, per exemple.

## **2.3. Descripció de les variables explicatives**

Les variables explicatives s'han dividit en diferents grups segons el tema a tractar, com són la data, la geografia, les substàncies involucrades, les que descriuen el tipus d'accident, les causes, la quantitat de substància involucrada, les que resumeixen l'accident i les complementàries.

### **2.3.1. Variables que indiquen la data de l'accident**

La data de l'accident es coneix a partir d'aquestes 3 variables:

- ❑ **DIA** és discreta i indica el dia en què va ocórrer l'accident. Quan no se sap el dia hi apareix espai en blanc (aquest és el cas de 372 accidents) ó 00 (com passa per 44 accidents). Per tant, es coneix la data de 8.374 accidents. El rang de valors és de 1 a 31.
- ❑ **MES** és discreta i indica el mes en què va ocórrer l'accident. Quan no se sap el mes apareix espai en blanc (aquest és el cas de 143 accidents) ó 00 (com passa per 19 accidents). Per tant, es coneix la data de 8.628 accidents. El rang de valors és de 1 a 12.
- ❑ **AÑO** és discreta i indica l'any en què va ocórrer l'accident. Quan l'accident va ocórrer en anys anteriors al 1.900, s'hi troba espai en blanc (aquest és el cas de 11 accidents) i si l'accident va ocórrer en el 1.900 trobem 00 (com passa per 6 accidents). El rang de valors és del 00 a 98.

Cal remarcar que tots els accidents que es van produir al 1.900 tenen com a dia i mes 01, a excepció de l'accident amb codi AN 2802 (aquesta variable s'explica en l'apartat 2.3.8) que té com a mes 00. Per aquest motiu, aquests es consideraran dades faltants.

### 2.3.2. Variables geogràfiques

- ❑ **PD** conté la densitat de població de l'àrea afectada per l'accident. Té 6.661 missings, és a dir, es coneix la densitat de població de l'àrea afectada per 2.129 accidents. Aquesta variable és qualitativa i té 3 categories:
  - T: Ciutat o alta densitat de població.
  - V: Poble o densitat mitjana de població.
  - R: Rural o densitat baixa de població.
- ❑ **LOCATION** situa la localitat on es va produir l'accident. Aquesta variable és categòrica i conté 343 missings, és a dir, es coneix la localitat de 8.447 accidents.
- ❑ **REGION** situa la regió on es va produir l'accident. Aquesta variable és categòrica i conté 2.624 missings, és a dir, es coneix la regió de 6.166 accidents.
- ❑ **COUNTRY** situa l'estat on es va produir l'accident. Aquesta variable és categòrica i conté 90 missings, és a dir, es coneix l'estat de 8.700 accidents.

Les variables **LOCATION** i **REGION** tenen moltes categories i per aquesta raó serà difícil fer-les servir en l'estudi. Només s'utilitzarà **COUNTRY**. Cal remarcar que aquesta variable no conté ben bé l'estat, ja que es troben categories com per exemple: SPAIN, TENERIFE i CANARY ISLAND. Com que **COUNTRY** pot ser interessant per l'estudi, s'haurà de solucionar aquest problema, ja que no és lògic que hi hagi zones diferents d'un mateix estat classificades per separat. També s'hauran de corregir errors de codificació, com per exemple, distingir ROMANIA de ROUMANIA.

### 2.3.3. Variables que indiquen les substàncies involucrades

- ❑ **MN** és categòrica i indica els noms de les substàncies que han intervingut en l'accident. En el cas de que hi hagin participat més d'una substància estan separades mitjançant una barra. Per exemple: NATURAL GAS/PROPANE/DIESEL FUEL. El nombre de categories per tant, és molt gran, 1.585 i segurament no es podrà fer servir en l'estudi. Conté 121 missings, és a dir, es disposa dels noms de les substàncies de 8.669 accidents.
- ❑ **MT** és categòrica i classifica el tipus de substància segons el seu estat físic en el moment de l'accident. Una substància pot pertànyer a una d'aquestes 11 categories:
  - DUST: Sòlid en forma de fines partícules d'una grandària de 100 micres que poden formar núvols de pols en l'aire.
  - SÒLID: Sòlid en general (no inclou pols).
  - LÍQUID: Líquid en general.
  - SOLUTION: Mescla de líquid/sòlid en el qual els components estan uniformement distribuïts.

- SLURRY: Suspensió de sòlid en líquid, amb partícules d'un grandària superior a 50 micres.
- SUPERHEATL: Producte líquid que manté el seu estat a una temperatura superior a la del seu punt d'ebullició degut a que està pressuritzat.
- GAS: Substància amb una temperatura crítica per sota de 50° C, o que a 50° C té una pressió de vapor major que 3 bar absoluts.
- RLGAS: Gas mantingut com a líquid mitjançant refrigeració, a pressió atmosfèrica o propera a ella.
- PLGAS: Gas a pressió atmosfèrica o pròxima a ella que es manté com a líquid sota pressió. (Pressió relativa > 500 mbar).
- DGAS: Gas dissolt sota pressió en un solvent, el qual pot ser absorbit en un material porós.
- LGAS: Gas mantingut com a líquid en circumstàncies no especificades.

En el cas de que hagin participat més d'una substància es troben els estats físics separats mitjançant una barra, encara que sigui el mateix estat. Per exemple: GAS/LÍQUID/LÍQUID. Aquest fet fa que hi hagi un total de 157 combinacions diferents, fet que dificulta l'estudi. Caldrà codificar-les de forma més general. Conté 1.504 missings, dels quals n'hi ha 44 on no es coneix l'estat de la primera substància que intervé en l'accident.

- **MH (Material Hazard)** és categòrica i registra les categories de perill de les substàncies que intervenen en l'accident. Està dividida en 8 categories:

- TO: indica que la substància és tòxica.
- FI: indica que la substància és inflamable.
- EX: indica que la substància és explosiva.
- CO: indica que la substància és corrosiva.
- RA: indica que la substància és radioactiva.
- CD: indica que la substància és refrigerada.
- AS: indica que la substància és asfixiant.
- OX: indica que la substància és oxidant.

Pot donar-se el cas de que la substància pertanyi a més d'una de les variables mencionades. Per exemple: TO-FI-, que vol dir que la substància o substàncies que han intervingut en l'accident són tòxiques i inflamables. Aquesta variable no té dades mancants.

- **MC** és el codi utilitzat per identificar la substància o la família a la que pertany, és per això que es pot trobar el mateix codi per diferents substàncies. Els codis utilitzats són els que apareixen en la publicació “Transport of Dangerous Goods. Recommendations of the Committee of Experts on the Transport of Dangerous Goods” 3<sup>rd</sup> edition, United Nations ST/SGAC10/1/REV3. Quan en un accident hi ha participat més d'una substància, els codis d'aquestes aniran separats per una barra. Per exemple: 9999/1114/9999 ó /1267 aquest últim exemple indica que en aquest accident han intervingut dues substàncies, però que es desconeix el codi de la primera substància. Cal fer constar que el codi 9999 vol dir que aquesta substància no pertany a una categoria específica, sinó a la categoria ALTRES. Aquesta variable numèrica té 85 missings dels quals en 7 casos no es coneix el codi de la primera substància, per tant, es disposa dels codis de les substàncies que han intervingut en 8.705 accidents.
  
- **NSUS** és discreta i indica el nombre de substàncies que han intervingut en l'accident. El rang de valors és de 1 a 9. En la majoria dels accidents hi ha intervingut una única substància. Aquest és el cas de 7.892 accidents. Després hi ha 629 accidents en els que hi ha intervingut 2 substàncies, 169 en els que hi ha intervingut 3 substàncies, 74 en els que hi ha intervingut 4 substàncies, 15 en els que hi ha intervingut 5 substàncies, 7 accidents en els que hi ha participat 6 substàncies, 2 accidents en els que hi ha participat 7 substàncies, 1 accident en el que hi ha participat 8 substàncies i 1 accident en el que hi ha participat 9 substàncies.



### 2.3.4. Variables que descriuen el tipus d'accident

- IT recull la informació sobre la forma en que es desenvolupa l'accident, mitjançant dos codis de forma seqüencial. Per exemple, un accident produït per la fuga d'un líquid que provoca una pèrdua de material i posteriorment un basal, estaria codificat de la següent manera: RELEASE-CONTREL; RELEASE-POOL, és a dir, DESCRIPCIÓ PRINCIPAL 1 – DESCRIPCIÓ SECUNDÀRIA 1; DESCRIPCIÓ PRINCIPAL 2 – DESCRIPCIÓ SECUNDÀRIA 2.

Conté 769 missings, és a dir, es coneix la forma en què es desenvolupen 8.021 accidents. Té les 15 categories següents:

- BLEVE: Trencament sobtat, d'un dipòsit o sistema que conté gas líquat pressuritzat, al contacte directe d'un foc.
- CONTREL: Fuita de material amb un cabal que es manté durant un període prolongat.
- DENSEXP: Explosió causada per la reacció química d'un material sòlid o líquid.
- DENSGSCL: Núvol de gas produït després de la fuga i més pesat que l'aire que l'envolta.
- DUSTEXP: Explosió d'una suspensió de pols en l'aire.
- EXPLODE: Alliberament d'energia produint gas a una temperatura, pressió i velocitat que poden causar danys al voltant.
- FIRE: Procés de combustió caracteritzat per calor, fum o flama, o qualsevol combinació d'aquests.
- GASCLD: Mescla aire/gas dins d'un particular embolcall de concentració límit (i densitat no definida).
- INSTREL: Fuita d'una quantitat de material durant un espai curt de temps (normalment pocs segons).
- MISTEXP: Explosió d'una dispersió de líquid en gas (diàmetre de partícules de líquid: 0.1E-6 a 5000E-6 metres).

- NEUGSCL: Núvol de gas que té una densitat igual a la de l'aire que l'envolta.
- PHYSEXP: Explosió en un sistema físic que no involucra necessàriament cap ignició. Per exemple, trencament d'un dipòsit a pressió. (No està inclòs BLEVE).
- POOL: Incendi on el combustible està en forma de basal en la base de l'incendi.
- RELEASE: Fuita.
- TORCH: Combustió de material que surt amb una inèrcia significativa d'un orifici.

Els descriptors principals són: EXPLODE, FIRE, GASCLD i RELEASE. Evidentment, els secundaris són la resta. Aquesta variable pot tenir un o dos descriptors principals i un o dos descriptors secundaris. El que més pot interessar és el descriptor principal i segurament es dividirà aquesta variable en dues.

- IS és categòrica i indica la font d'ignició de l'accident. Pot tenir fins a dos codis per descriure-la. Té 6.385 missings, és a dir, es coneix la font d'ignició de 2.405 accidents. Té les 17 categories següents:

- AUTOIGNITE: Autoignició.
- ELECTRIC-DOMESTICS: Espurnes elèctriques en dispositius domèstics.
- ELECTRIC-INSTRUMENT: Espurnes elèctriques en instruments.
- ELECTRIC-MOTorgen: Espurnes elèctriques en motors elèctrics/generadors.
- ELECTRIC-STATIC: Espurnes elèctriques estàtiques (inclou enllumenat).
- ELECTRIC-VEHICLE: Espurnes elèctriques en vehicles elèctrics.
- FLAMES-BOILER: Flama lliure en calderes.
- FLAMES-FLARE: Flama lliure en flames.
- FLAMES-MATCH: Flama lliure en mixtos.
- FLAMES-STOVE: Flama lliure en forns/estufes.
- FRICSPARK-MACHINERY: Espurnes per fregament en maquinària rotativa.

- FRICSPARK-SPARKTOOL: Espurnes per fregament en eines que produeixen espurnes.
  - HOTSURFACE-FRICSURF: Superfície calent per efecte de fricció.
  - HOTSURFACE-INCAND: Superfície calent per partícules incandescentes.
  - HOTSURFACE-STEAMPIPE: Superfície calent per canonades de vapor (línies de procés).
  - HOTSURFACE-VEXHAUSTS: Superfície calent per escapament de vehicles.
  - NONIGNITE: No ignició de materials inflamables.
- **IG** és discreta i indica el temps, en segons, que transcorre des de que s'origina l'incident desencadenant fins l'explosió. S'estima a partir de la distància a la font d'ignició i la quantitat de substància que intervé. El seu rang de valors és de 0 a 18.000 segons. Té 8.632 missings, la qual cosa vol dir que només es coneix l'estimació dels temps d'ignició de 158 accidents. Això fa que no es pugui aprofitar per l'estudi.

### 2.3.5. Variables que expliquen les causes de l'accident

- **OG** indica l'origen de l'accident, mitjançant dos codis. Per exemple, un accident produït en la planta d'emmagatzematge, i més concretament en un dipòsit d'emmagatzematge a pressió atmosfèrica, estaria codificat de la següent manera: STORAGE; ASVESSEL, és a dir, TIPUS D'ACTIVITAT PRINCIPAL; ORIGEN ESPECÍFIC.

Conté 187 missings, és a dir, es coneix l'origen de 8.603 accidents. Té 29 categories i són les següents:

- ASVESSEL: Dipòsits a pressió atmosfèrica utilitzats per l'emmagatzematge de productes i matèries primeres.
- BARGE: Dipòsit transportat per aigua terra endins (com riu).
- COMMTANK: Dipòsit comercial petit.
- DOM/COM: Accident originat en locals domèstics o comercials.
- FIREDEQUIP: Equips de procés amb flama, inclou forns, incineradors, piles i xemeneies.
- HEATXCHANG: Intercanviadors de calor, inclou carcassa i tub. Evaporadors de plats, condensadors, calderes i recalentadors.
- HOSE: Mànegues i altres connexions per càrrega/descàrrega i similars.
- MACDRIVE: Dipòsits de procés, inclou monitors elèctrics, motors de combustió i turbines.
- PACKAGE: Contenidors de transport portàtil, inclou tambors, tonells, bidons, caixes, bosses, paquets compostos i cilindres.
- PIPELINE: Canonades exteriors utilitzades pel transport a granel extern a la planta.
- PIPEWORK: Canonades de planta, vàlvules i juntes associades.
- PROCESS: Accident originat en unitats o àrees d'una planta de procés.
- PSVESSEL: Dipòsits pressuritzats utilitzats per l'emmagatzematge de productes i matèries primeres.
- PUMP: Qualsevol tipus de bomba, compressor, ejector o ventilador.
- PVESSEL: Dipòsits de procés que inclou unitats com centrifugadores, torres, columnes, assecadors, destil·ladors, absorbidors, filtres, ciclons, intercanviadors iònics, cristal·litzadors, etc.
- RAILTANKER: Dipòsits pressuritzats de propòsit general, del tipus "Tank Cars" USA.
- REACTOR Reactor de procés. Inclou continu, discontinu, tubular, torre, etc.
- ROADTANKER: Dipòsits simples, compartimentats o múltiples, de tipus "Tank Trucks" USA.
- SHIP: Dipòsit que va per l'oceà.
- SIZECHANGE: Equip per la reducció/ampliació de la grandària, inclou molins, afiladors, premses, trencadors, talladors i aglomeradors.

- SOLIDMOVE: Equip per moure material sòlid, per exemple: transportador, cinta, elevadors, paleta, cargols i pneumàtics.
- SOLIDSTORE: Emmagatzematge de sòlids, inclou arços, sils i tonells.
- STORAGE: Accident originat en unitats o àrees d'una planta d'emmagatzematge.
- SUBSTATION: Subestació.
- TANKCONTNR: Tancs amb capacitat igual a 50 litres que té un embolcall exterior dotat d'instrumentació de servei i equips estructurals. Capaç de ser transportat per terra o mar i ser carregat/descarregat sense necessitats de retirar l'equipament estructural. Tenen elements estabilitzadors externs en l'embolcall que poden ser elevats quan estan plens.
- TRANSFER: Accident produït durant operacions de càrrega i descàrrega.
- TRANSPORT: Accident originat durant el transport de material externament a la planta. Està inclòs canonades (PIPELINES).
- WAREHOUSE: Accident originat en un magatzem de mercaderies.
- WASTE: Accident originat en un magatzem o àrea de residus incloent pous de contenció, abocadors, contenidors de residus a granel. S'exclouen materials que s'estan utilitzant en planta de producció.

Els tipus d'activitat són: DOM/COM, PROCESS, STORAGE, TRANSFER, TRANSPORT, WAREHOUSE i WASTE. Els orígens específics són la resta. Aquesta variable té com a màxim un tipus d'activitat i un origen específic. El que més interessa és el tipus d'activitat, ja que es disposen de moltes categories específiques i segurament es dividirà aquesta variable en dues.

- **GC** és categòrica i fa referència a la causa general que va desencadenar l'accident. Té 3.451 missings, per tant, es coneix la causa general de 5.339 accidents. Aquestes variables té les següents 8 categories:

- EXTERNAL: Esdeveniments externs.
- HUMAN: Factor humà.
- IMPACT: Fallida per impacte.

- INSTRUMENT: Fallida d'instrumentació.
- MECHANICAL: Fallida mecànica.
- PROCOND: Variació de les condicions de procés.
- VREACTION: Reacció violenta.
- SERVICE: Fallida de serveis.

Tot i que semblen poques categories, com que hi ha la possibilitat de què intervingui més d'una causa, es pot trobar combinacions. Exemple: VREACTION-HUMAN. La categoria EXTERNAL té una variant que és EXTERNAL' que correspon a la mateixa categoria, EXTERNAL, ja que es va produir un error en passar les dades de la base original a la que es disposa per l'estudi. S'han trobat 13 errors que a més tenen en comú la categoria SABOTAGE a la variable SC. Això s'haurà d'arreglar, fent que totes dues versions siguin només una categoria.

- SC és categòrica i descriu les causes específiques de l'accident. Té 4.192 missing, és a dir, es coneixen les causes específiques de 4.598 accidents. Aquesta variable té les 56 categories següents:

- ACCVENT: Fallida per un factor humà deguda a un ventament accidental.
- ALARM: Fallida d'instruments (alarma).
- BRITTLE: Fallida mecànica deguda a una fallida fràgil.
- COMMUNICAT: Fallida per un factor humà deguda a comunicacions.
- COMPAIR: Fallida de serveis (aire/nitrogen comprimit).
- COMPUTER: Fallida de l'ordinador.
- CONEXP: Fallida per explosió confinada.
- CONNECTING: Fallida deguda a la connexió.
- CONSTRUCT: Fallida per un factor humà deguda a errors en la construcció.
- CONTROL: Fallida d'instruments (controlador).
- CORRODE: Fallida mecànica per corrosió.
- CRANE: Fallida per impacte d'una grua.

- DESIGN: Fallida per un factor humà deguda a errors de disseny.
- DRAINACC: Fallida per un factor humà deguda a un accident en el drenatge.
- EARTHQUAKE: Fallida per un terratrèmol.
- ELECTRIC: Fallida de serveis (electricitat).
- EXCAVEQUIP: Fallida per impacte d'un equip d'excavació.
- EXTNLEXP: Fallida per explosió.
- EXTNLFIRE: Fallida per incendi.
- FATIGE: Fallida mecànica per fatiga.
- FLOODS: Fallida per inundacions.
- GAS: Fallida de serveis (gas).
- GENERALOP: Fallida per un factor humà deguda a operacions generals.
- GLANDSEAL: Fallida en un segell de prensaestopes.
- GROUND: Fallida per un enfonsament, fatiga del terreny o erosió del suport.
- HIGHWINDS: Fallida per vents forts.
- HOSE: Fallida mecànica deguda a la mànega.
- HVYOBJECT: Fallida per impacte d'un objecte pesat.
- INCOMPAT: Fallida mecànica per ús de materials incompatibles.
- INDICATOR: Fallida d'instruments (indicador).
- INSTALL: Fallida per un factor humà deguda a errors en la instal·lació.
- INTNLFIRE: Fallida per foc intern.
- ISOLUNCUP: Fallida per aïllament/denatge abans del descarrilament.
- LIGHTNING: Fallida per llampecs.
- MAINTAIN: Fallida per un factor humà deguda a un manteniment general.
- MANAGEMENT: Fallida per un factor humà deguda a un error de direcció general.
- METALLURG: Fallida mecànica per altres errades metal·lúrgiques.
- OVERFILL: Fallida per un factor humà deguda a un sobreple.
- OVERHEAT: Fallida mecànica per sobreescalfament.
- OVERLOAD: Fallida mecànica per sobrecàrrega.
- OVERPRES: Fallida mecànica per sobrepressió.
- PROCEDURES: Fallida per un factor humà deguda a procediments.

- RAILACC: Fallida per accident ferroviari en el que no estan involucrats altres vehicles.
- RELIEFVALV: Fallida mecànica deguda a la vàlvula d'alleujament.
- ROADACC: Fallida per accident en carretera en el qual no estan involucrats altres vehicles.
- RUNAWAY: Fallida per reacció fora de control.
- SABOTAGE: Fallida per sabotatge/vandalisme.
- SHIP/LAND: Fallida per comissió entre un vaixell i terra.
- SHIP/SHIP: Fallida per col·lisió entre vaixells.
- STEAM: Fallida de serveis (vapor).
- TEMPRTURE: Fallida per temperatures externes.
- TRIP: Fallida d'instruments (trinquet).
- VALVE: Fallida deguda a una vàlvula.
- VEHICLE: Fallida per impacte amb un altre vehicle.
- WATER: Fallida de serveis (subministrament d'aigua).
- WELDFAIL: Fallida mecànica deguda a la soldadura.

Pot donar-se el cas de que un accident tingui més d'una causa específica i per tant, poden haver-hi combinacions de les categories mencionades. Per exemple: CONNECTING-DESIGN, que vol dir que l'accident s'ha produït degut a un error en la connexió que s'hauria pogut evitar amb un disseny millor.



### 2.3.6. Variable que mesura la quantitat de substància involucrada

- **QY** és una variable de text que està formada per nombres, però alguns contenen símbols matemàtics, com per exemple: <1000. Estima la quantitat de substància intervinguda en l'accident en tones. Té 6.497 missings, és a dir, que només es coneix l'estimació de la quantitat de substància que ha intervingut en 2.293 accidents. Aquesta variable també té el problema que conté símbols de desigualtats que dificultarà la seva transformació a variable numèrica. En el fet de que hi aparegui la quantitat 99999, vol dir que és la quantitat mínima de substància intervinguda en l'accident, ja que MHIDAS té com a màxim 99999.

### 2.3.7. Variables que expliquen l'accident

Aquestes variables no es faran servir per l'estudi, ja que no es poden modelar al ser variables de text:

- **AB** fa un breu resum de l'accident, amb informació més o menys detallada. Només té dos missings (per tant, es disposa de 8.788 observacions) que correspon als accidents amb codi AN 5637 i 721 (aquesta variable s'explica en l'apartat 2.3.8).
- **KW** indica a través d'unes paraules clau acompanyades d'un text si existeix més informació sobre alguns aspectes addicionals. Per exemple: AFTER, vol dir que existeix informació addicional sobre els efectes que l'accident ha provocat posteriorment. Té 2.328 missings, és a dir, es tenen les paraules clau de 6.462 accidents.

### 2.3.8. Variables complementàries

- **AN** és discreta i indica el codi de l'accident. Analitzant aquesta variable s'hi han trobat registres diferents amb el mateix codi. Això és un error a l'introduir les dades o del pas de la base de dades de substàncies a la d'accidents i per tant s'haurà de corregir. Aquests codis són: 1023, 1639, 179, 1941, 2122, 2152, 2322, 2446, 2483, 289, 2901, 3151, 323, 398, 6232, 6699, 675, 7144, 745, 749 i 7717. Aquesta variable no es farà servir per la construcció de models.
- **CR** indica la font d'informació de l'accident. Hi ha 42 missings, és a dir, es disposa de 8.748 accidents on es coneix la font d'informació. Aquesta és una variable qualitativa amb aquestes 9 categories:
  - AEA
  - DURRAN
  - HSE: Health and Safety Executive
  - ICHIME
  - KAAT
  - OECD: Organització de Cooperació i Desenvolupament Econòmic.
  - PAHO
  - SRD: Safety and Reliability Directorate
  - WORKSA

De les sigles que no s'indica el significat, es perquè aquest no es coneix.

- **RA** és de text i indica el nombre d'articles o textos disponibles sobre l'accident que es poden consultar. Aquesta variable està formada per nombres, però alguns contenen el símbol de desigualtat >, com per exemple: >60. Té 27 missings, i per tant, es disposa dels textos sobre 8.763 accidents.

## 2.4. Resum del més important

De totes les variables que conté la base de dades, parlant amb l'enginyer responsable del projecte, es va considerar que les variables més interessants com a variable resposta són: **KILLED** (nombre de morts en l'accident), **INJURED** (nombre de persones ferides en l'accident), **EVACUATED** (nombre de persones evacuades en l'accident) i **DAMAGE** (estimació econòmica del dany material), encara que aquesta última variable té poques observacions. Totes aquestes variables més la variable **QY** (estimació de la quantitat de substància involucrada en l'accident) tenen el problema de que són de text, degut als símbols matemàtics de desigualtats ( $>$  o  $<$ ) que contenen. Això serà un dels problemes que s'hauran de resoldre inicialment. A més a més, les variables resposta tenen molts zeros, i això dificultarà el seu modelat.

Les variables explicatives potencialment més importants són: **AÑO**, **NSUS** (nombre de substàncies que han intervingut en l'accident), **MT** (estat físic de la substància que ha participat en l'accident), **IT** (tipus d'accident), **OG** (origen de l'accident), **GC** (causa de l'accident), **COUNTRY** (estat on ha succeït l'accident) i **QY** (quantitat de substància involucrada en l'accident). Les variables **PD** (densitat de la població), **IS** (font d'ignició) i **IG** (temps que transcorre des de que s'origina l'incident desencadenant fins l'explosió o incendi) també són potencialment importants, però no es faran servir per falta d'observacions.

Aquesta base de dades presenta una sèrie de problemes: cal remarcar que hi ha accidents diferents amb el mateix codi d'identificació (**AN**), que tenen característiques comuns i algunes diferències. La variable **COUNTRY** que hauria de classificar els accidents segons l'estat on es van produir, presenta noms que no són estats, estats repetits, però considerats diferents per alguna errada en la codificació i problemes semblants.

Hi ha variables que tenen moltes categories, com per exemple, **IT** i **OG**. La qual cosa portarà a fer una reclassificació, d'acord amb l'enginyer responsable del projecte, de forma que s'agrupin les categories més semblants.

També existeix el problema de que hi ha variables que contenen molts missings, com per exemple, **IG**, i per aquest motiu hi haurà variables que no es podran fer servir per no disposar de suficients dades per treballar-hi. A més a més, no totes les variables tenen la mateixa codificació pel que fa als missings, la qual cosa comporta confusions, ja que a vegades estan codificats amb valors possibles, com per exemple el zero en la variable **DAMAGE**.

A continuació es mostra la taula de totes les variables que componen la base de dades MHIDAS. Cal recordar que les variables **KILLED**, **INJURED**, **EVACUATED**, **DAMAGE** i **QY** es transformaran a numèriques. El nombre total d'observacions és 8.790.

| VARIABLE                            | CATEGORIES | RANG     | ZEROS | MISSINGS           | NO MISSINGS |
|-------------------------------------|------------|----------|-------|--------------------|-------------|
| KILLED                              | TEXT       | -        | 5.342 | 1.846              | 6.944       |
| INJURED                             | TEXT       | -        | 5.038 | 986                | 7.804       |
| EVACUATED                           | TEXT       | -        | 4.084 | 3.293              | 5.497       |
| DAMAGE                              | TEXT       | -        | -     | 7.850 <sup>1</sup> | 940         |
| DIA                                 | NUMÈRICA   | 1-31     | -     | 416                | 8.374       |
| MES                                 | NUMÈRICA   | 1-12     | -     | 162                | 8.628       |
| AÑO                                 | NUMÈRICA   | 0-98     | -     | 0                  | 8.790       |
| PD (densitat de població)           | 3          | -        | -     | 6.661              | 2.129       |
| LOCATION                            | 4.711      | -        | -     | 343                | 8.447       |
| REGION                              | 697        | -        | -     | 2.624              | 6.166       |
| COUNTRY                             | 165        | -        | -     | 91                 | 8.699       |
| MN (nom de la substància)           | 2.095*     | -        | -     | 121                | 8.669       |
| MT (estat físic de la substància)   | 11<br>156* | -        | -     | 1.504              | 7.286       |
| MH (ris de la substància)           | 8<br>24*   | -        | -     | 0                  | 8.790       |
| MC (codi de la substància)          | 1.008*     | -        | -     | 85                 | 8.705       |
| NSUS (nombre de substàncies)        | NUMÈRICA   | 1-9      | -     | 0                  | 8.790       |
| IT (tipus d'accident)               | 15         | -        | -     | 769                | 8.021       |
| IS (font d'ignició)                 | 17         | -        | -     | 6.385              | 2.405       |
| IG (temps fins l'explosió)          | NUMÈRICA   | 0-21.600 | -     | 8.632 <sup>1</sup> | 158         |
| OG (origen de l'accident)           | 29         | -        | -     | 187                | 8.603       |
| GC (causa general de l'accident)    | 8<br>139*  | -        | -     | 3.451              | 5.339       |
| SC (causa específica de l'accident) | 56<br>579* | -        | -     | 4.192              | 4.598       |
| QY (quantitat de substància)        | TEXT       | -        | -     | 6.497 <sup>1</sup> | 2.293       |
| AB (resum accident)                 | TEXT       | -        | -     | 2                  | 8.788       |
| KW (text amb paraules clau)         | TEXT       | -        | -     | 2.328              | 6.462       |
| AN (codi accident)                  | NUMÈRICA   | 1-9.003  | -     | 0                  | 8.790       |
| CR (font informació)                | 9          | -        | -     | 42                 | 8.748       |
| RA (nombre articles)                | TEXT       | -        | -     | 27                 | 8.763       |

Taula 1: descripció de les variables de la base de dades MHIDAS.

<sup>1</sup> Codificats amb 0

\* nombre de combinacions.

# CAPÍTOL 3: ADAPTACIÓ DE LA BASE DE DADES MHIDAS

## 3.1. Introducció

En aquest capítol es presenten les modificacions realitzades sobre la base de dades original amb l'objectiu d'ajustar-la en tot el que sigui necessari a l'estudi a realitzar. S'ha decidit amb l'enginyer responsable del projecte fer una *restricció de l'àmbit d'estudi*, tot limitant l'estudi als accidents produïts en instal·lacions fixes. Per aquesta restricció es necessària la variable **OG** que indica el tipus d'activitat industrial en el moment de l'accident, i la instal·lació on es va originar. Això es fa a través de dos codis, el primer conté el tipus d'activitat i el segon l'origen.

Els tipus d'activitats són: DOM/COM, PROCESS, STORAGE, TRANSFER, TRANSPORT i WAREHOUSE. D'acord amb l'enginyer responsable del projecte, s'han tret de la base de dades els accidents que estan a la categoria TRANSPORT de la variable **OG**. Així s'exclouen els accidents derivats durant feines de transport, ja sigui per terra o mar. Després de fer servir aquest filtre, la base de dades ha quedat reduïda a 5.176 accidents amb substàncies químiques perilloses.

## **3.2. Depuració de les variables resposta**

### **3.2.1. Depuració de les variables nombre de morts, nombre de ferits i nombre d'evacuats**

Les variables **KILLED**, **INJURED** i **EVACUATED** s'han transformat a numèriques. Aquestes variables contenen valors o estimacions a l'alça o a la baixa, com per exemple, en evacuats: > 200 ó en ferits: <10. Transformar a numèriques aquestes variables de text que tenen estimacions no és senzill. Una de les opcions preses ha sigut prescindir dels símbols de desigualtat. Per exemple: > 200, ara és 200 i <10 és 10, ja que la informació disponible no permet afinar més.

Un altre problema que presenten aquestes variables és que tenen zeros que corresponen a dades mancants i per tant, en comparar la base de dades original (de substàncies) amb la que es disposa per l'estudi (d'accidents) s'ha detectat que hi havien més zeros dels que realment havien de ser-hi. El que s'ha fet ha sigut arreglar aquest problema posant com a 999999 els valors mancants que estaven codificats com a espai en blanc en la base de dades original (de substàncies), i com a zero a la base de dades adaptada. Per poder dur a terme aquesta modificació s'ha utilitzat consultes d'actualització mitjançant el programa ACCESS. El que ha obligat a profunditzar els coneixements en aquest programa.

Cal recordar que els codis dels accidents en la base de dades adaptada es poden relacionar amb els de la base de dades de substàncies. L'única diferència és que en la base de dades de substàncies hi ha el codi de l'accident més una lletra, segons les substàncies que hagin participat en l'accident. Per exemple: si han intervingut dues substàncies hi apareix, el codi més la lletra A i en un altre registre (o fila) el codi més la lletra B, (7250 A i 7250 B). En la base de dades adaptada, la d'accidents hi ha només el codi 7250 i la variable **NSUS** indica el nombre de substàncies que han intervingut en l'accident, en aquest cas dues. Hi han intervingut tantes substàncies com lletres apareguin.

Com que treballarem amb el programa MINITAB, després els 999999 han estat transformats en asterisc “\*”, ja que és el símbol que aquest programa utilitza per a identificar les dades mancants.

Després de transformar aquestes variables a numèriques, s’ha creat una columna al costat indicant com estava aquesta variable originalment, codificada amb 1, 2, 3. Si en la base de dades original hi ha “un menor que” a la nova variable hi apareix un 1, si hi ha “un major que”, a la nova variable hi apareix un 2, i si hi ha un valor concret hi apareix un 3. D’aquesta manera, si més endavant es decideix fer una transformació més sofisticada, aquesta variable proporcionarà la informació necessària, i no s’haurà perdut informació.

Un cop creades aquestes variables, s’ha observat que no coincidia el nombre de missings amb els de la variable sense els símbols matemàtics. Això es degut a que en la base de dades de substàncies hi ha registres on es coneix el nombre de persones mortes, ferides i evacuades i en una altra substància del mateix accident no. Aquests accidents tenen els següents codis AN: 558, 3097, 3117, 3182, 3683, 1650, 1013, 7044 i 7144. A continuació s’especifica com està recollida la informació en la base de dades original de les substàncies que han intervingut en aquests accidents i la solució adoptada en cada cas.

En l’accident que té per codi AN 558 hi ha participat dues substàncies. En la fila on es registra una d’elles hi ha zero persones mortes i ferides, i missing en evacuades, i en l’altra fila, on es registra l’altre substància, no es coneix ni el nombre de persones mortes, ni el de ferides ni el d’evacuades. S’ha deixat el nombre que es coneix de persones mortes i ferides perquè es considera que és millor disposar d’una resposta coneguda abans d’una dada mancant.



En l'accident que té per codi AN 3097 hi ha participat dues substàncies. En la fila on es registra una d'elles hi ha zero persones mortes, una persona ferida i 200 persones evacuades, i en l'altra fila on es registra l'altre substància no es coneix ni el nombre de persones mortes, ni el d'evacuades, però es coneix que hi ha hagut una persona ferida. El que s'ha fet ha estat deixar zero persones mortes, una ferida i 200 evacuades.

En l'accident que té per codi AN 3117 hi ha participat dues substàncies. En la fila on es registra una d'elles no hi ha cap persona morta ni ferida, i 30.000 persones evacuades, i en l'altra fila on es registra l'altre substància no es coneix ni el nombre de persones mortes, ni el de ferits, però es coneix que han hagut 30.000 evacuats. El que s'ha fet ha estat deixar zero persones mortes i ferides i 30.000 evacuades.

En l'accident que té per codi AN 3182 hi ha participat dues substàncies. En la fila on es registra una d'elles no hi ha cap mort ni ferit i 800 evacuats, i en l'altra fila on es registra l'altre substància no es coneix res sobre les tres variables mencionades. El que s'ha fet ha estat deixar la informació que es coneix.

En l'accident que té per codi AN 3683 hi ha participat dues substàncies. En la fila on es registra una d'elles hi ha zero persones mortes i ferides i no es coneix el nombre d'evacuats, i en l'altra fila on es registra l'altre substància no es coneix res sobre les variables resposta. El que s'ha fet ha estat deixar zero persones mortes, zero ferides i missing en evacuats.

En l'accident que té per codi AN 1650 hi ha participat dues substàncies. En la fila on es registra una d'elles hi ha 5 persones mortes, 25 ferides i 1.000 evacuades, i en l'altra fila on es registra l'altre substància no es coneix res sobre les variables resposta. El que s'ha fet ha estat deixar les dades que es coneixen.

En l'accident que té per codi AN 1013 hi ha participat tres substàncies. En dues d'elles hi ha zero persones mortes i ferides i no es coneix el nombre de persones evacuades, i en l'altra substància no es coneix res sobre les variables resposta. El que s'ha fet ha estat deixar les dades que es coneixen.

En l'accident que té per codi AN 7044 hi ha participat tres substàncies. En dues d'elles hi ha zero persones ferides i no es coneix ni el nombre de persones evacuades ni el nombre de morts, i en l'altra substància no es coneix res sobre les variables resposta. El que s'ha fet ha estat deixar les dades que es coneixen.

En l'accident que té per codi AN el 7144 hi ha participat quatre substàncies. En tres d'elles no es coneix el nombre de morts, hi ha menys de 70 persones ferides i 12.000 evacuats, i en l'altra substància no es coneix res sobre les variables resposta. El que s'ha fet ha estat deixar les dades que es coneixen.

Un cop depurats els errors, el contingut de les variables resposta és:

- **KILLED** és discreta i indica el nombre de morts en l'accident. Conté 3.024 missings codificats amb el nombre 999999, per tant, es coneix el nombre de morts de 2.152 accidents. El seu rang de valors és de 0 a 2.000.
- **INJURED** és discreta i conté el nombre de persones ferides en l'accident. Conté 2.374 missings codificats amb el nombre 999999, per tant, es coneix el nombre de ferits de 2.802 accidents. El seu rang de valors és de 0 a 9.999.
- **EVACUATED** és discreta i indica el nombre de persones evacuades en l'accident. Conté 4.055 missings codificats amb el nombre 999999, per tant, es coneix el nombre d'evacuats de 1121 accidents. El seu rang de valors és de 0 a 200.000.

### 3.2.2. Depuració de la variable danys econòmics

La variable **DAMAGE** s'ha transformat a numèrica i s'ha anomenat **DAMAGE1**. Per fer que aquesta transformació, s'han tret els símbols de desigualtat matemàtics i s'ha deixat el valor màxim o mínim de l'estimació econòmica del dany material. Els nombres que no contenien cap símbol de desigualtat, s'han deixat igual. Cal dir que s'ha creat una columna al costat indicant com estava aquesta variable originalment, codificada amb 1, 2, 3, tal i com s'ha fet amb les variables resposta comentades anteriorment. D'aquesta manera, si més endavant es decideix fer una transformació més sofisticada, aquesta variable proporcionarà la informació necessària i no s'haurà perdut informació.

- **DAMAGE1** és continua i fa una estimació econòmica del dany material provocat per l'accident en milions de dòlars. En comparar la base de dades de substàncies amb l'adaptada s'han trobat 4 accidents amb el dany econòmic igual a zero que s'havien confós amb les dades mancants, ja que els missings també estaven codificats com 0 originalment. Aquests 4 casos tenen com a AN: 992, 1181, 1752 i 432. Ara els missings s'han codificats com a "\*" per evitar confusions. Té 4.409 missings, per tant, només es disposa de 767 accidents on es coneix l'estimació econòmica del dany material. El seu rang de valors és de 0 a 370 milions de dòlars.

Per poder fer comparables les xifres de la variable anterior entre els diferents anys s'han creat les següents variables:

- **DAMAGE\_CEI** és continua i conté la valoració econòmica del dany, actualitzada a l'any 1.999 mitjançant l'índex d'actualització del cost CEI (Chemical Engineering Cost Index). Aquest índex fa les actualitzacions a partir de l'any 1.947, és a dir, per anys anteriors a aquest no es té actualització. Té 4.433 missings codificats mitjançant un "\*", és a dir, s'hi pot treballar amb 743 observacions. El rang de valors és de 0,002 a 404,913 milions de \$.

- **DAMAGE\_MAR** és continua i conté la valoració econòmica del dany material, actualitzada a l'any 1.999 mitjançant l'índex d'actualització del cost Marshall & Stevens. Aquest índex fa les actualitzacions a partir de l'any 1.913, és a dir, per anys anteriors a aquest no es té actualització. Té 4.417 missings, codificats amb un "\*", per tant hi ha 459 estimacions econòmiques actualitzades. El rang de valors és del 0,002 a 456,533 milions de \$.
  
- **ÍNDEX\_CEI** és continua i conté l'índex d'actualització de cost CEI. No té cap missing, ja que es coneixen tots els índexs CEI a partir de l'any 1.947, si l'any és anterior a 1.947, el valor és zero. El rang de valors és del 0 a 389,5.
  
- **ÍNDEX\_MAR** és continua i conté l'índex d'actualització de cost Marshall & Stevens. No té cap missing, ja que es coneixen tots els índexs Marshall a partir de l'any 1.913. Si l'any és anterior al 1.913, el seu valor és zero. El rang de valors és de 0 a 1.061,9.

A partir dels índexs CEI i Marshall & Stevens, i coneixent l'any d'ocurrència de l'accident s'ha calculat l'actualització econòmica del dany material mitjançant la següent fórmula:

$$Valor\_actual(1999) = Valor\_any(X) * \frac{Índex\_any(1999)}{Índex\_any(X)}$$

De totes aquestes variables la que es farà servir per estudiar el dany material en aquest estudi és **DAMAGE\_MAR**, ja que és la variable que té més observacions actualitzades (a partir d'ara s'anomenarà **DAMAGE**).

### **3.3. Depuració de les variables explicatives**

#### **3.3.1. Depuració de la data de l'accident**

- **DIA** i **MES** contenen el dia i el mes en què es va produir l'accident, respectivament. En la base de dades original, els missings estan codificats com a espais en blanc ó 00. Per tant, els 00 i els espais en blanc s'han transformat en "\*", ja que el programa MINITAB no reconeix com a missings els 00 i si els "\*". Els accidents que es van produir l'any 1.900 tenen com a dia i mes 01 a excepció de l'accident amb AN 2802 que té com a mes 00. D'acord amb l'enginyer responsable del projecte s'ha decidit considerar-los missings. La variable **DIA**, conté 307 missings, per tant, es disposa del dia en 4869 accidents. El rang de valors és de 1 a 31, ambdós inclosos. La variable **MES** conté 141 missings (es disposa del mes de 5.035 accidents). El rang de valors és de 1 a 12, ambdós inclosos.

La variable **AÑO** de la base de dades original a partir d'ara s'anomenarà **ANY**.

- **ANY** indica l'any en què es va produir l'accident. Els accidents anteriors a 1.900 es codifiquen com a 00 i els que van succeir al 1.900 també, ja que no interessa gaire diferenciar aquests accidents perquè són molt antics i no són representatius dels actuals o dels futurs. Aquesta variable no conté cap missing. El rang de valors va de 00 al 98.

S'ha decidit construir una nova variable indicadora per diferenciar els accidents produïts abans de la dècada dels setanta, codificats mitjançant un zero, dels que es van produir amb posterioritat, codificats mitjançant un 1. Donat que la informació dels accidents de la base de dades MHIDAS es va començar a recopilar a començaments dels anys vuitanta, la informació dels accidents ocorreguts abans del setanta s'ha recollit de registres d'hemeroteques, i només es recullen els accidents més greus, amb poca informació i de poca qualitat. Dels accidents produïts després dels anys 70, en canvi es té més informació i de més qualitat, encara que siguin menys greus. Aquesta variable s'ha anomenat **ANTIGUITAT**.

També s'ha cregut oportú crear una variable que indiqui la dècada en què es va produir l'accident amb el nom **DÈCADA**. Aquesta variable té un 1 pel període d'anys comprès entre 1.910 i 1.919, la dècada dels anys 10, té un 2 pel període d'anys comprès entre 1.920 i 1.929, la dècada dels anys 20, i així successivament fins arribar a la dècada dels 90, codificats amb un 9. Els accidents produïts abans del 1.900 i els produïts des de 1.900 a 1.909 (només hi ha accidents d'aquest interval produïts a l'any 1.901) estan codificats amb un 0 per tal de poder-los diferenciar de la resta. Per tant, de totes les variables que indiquen la data de l'accident només utilitzarem aquestes dues últimes variables (**ANTIGUITAT** i **DÈCADA**).

### 3.3.2. Depuració de les variables geogràfiques

- ❑ **PD** indica la densitat de població de l'àrea afectada per l'accident. Conté 3.928 missings, és a dir, es coneix la densitat de població de l'àrea afectada per 1.248 accidents. Aquesta variable és qualitativa i té aquestes 3 categories: T, V i R.
- ❑ **LOCATION** és la localitat on va passar l'accident. Conté 193 missings. Es disposa de la localitat de 4.983 accidents.

- **REGION** és la regió on va succeir l'accident. Conté 1.560 missings, és a dir, es coneixen les regions on es van produir 3.616 accidents. De totes maneres si es coneix la localitat s'hi pot saber a quina regió pertany l'accident, i el nombre de missings disminuiria fins a 193.

Tant la variable **LOCATION** com **REGION** tenen moltes categories, per aquest motiu no s'utilitzaran en l'estudi.

- **COUNTRY** indica el país on va passar l'accident. Conté 32 missings. Es coneix el país de 5.144 accidents. Cal recordar que aquesta variable no és ben bé el país, ja que es troben categories com SPAIN, TENERIFE i CANARY ISLAND i països codificats amb noms diferents, com ROMANIA i ROUMANIA etc. Però com no s'utilitzarà (ja que farem servir una variable geogràfica més global), de moment no se solucionarà aquest problema.

Per tal de fer una classificació geogràfica més general, s'ha creat la variable **MÓN**:

- **MÓN** és categòrica i classifica els estats en funció del seu grau de desenvolupament tecnològic. L'assignació dels estats s'ha fet en funció de la seva renda per càpita, encara que en alguns casos no s'ha seguit aquest criteri, ja que una renda mitjana elevada no vol dir necessàriament un nivell alt de desenvolupament tecnològic. No conté missings i té tres categories:

- 1: Estats desenvolupats.
- 2: Estats en via de desenvolupament.
- 3: Estats subdesenvolupats.

A partir d'aquestes categories s'han creat 3 variables indicadores amb els noms **MDESENV**, **MENVIADSENV** i **MSUBDESENV**.

A continuació es mostra la classificació dels estats segons aquesta categorització, es poden observar alguns errors ja mencionats:

| MÓN                                 | ESTAT  |  |  |
|-------------------------------------|--|--|--|
| ESTATS<br>DESENVOLUPATS             | AUSTRALIA<br>AUSTRIA<br>BELGIUM<br>CANADA<br>CANARY ISLANDS<br>DENMARK<br>EIRE<br>ENGLISH<br>FINLAND   |  |  |
| ESTATS EN VIA DE<br>DESENVOLUPAMENT | ARGENTINA<br>BRAZIL<br>BULGARIA<br>CYPRUS<br>CZECH REPUBLIC<br>CZECHOSLOVAKI<br>CHILE<br>CHINA<br>ESTONIA  |  |  |
| ESTATS<br>SUBDESENVOLUPATS          | AFGHANISTAN<br>AFRICA<br>ALBANIA<br>ALGERIA<br>ANGOLA<br>ANTARCTICA<br>ANTILLES<br>ARABIAN GULF<br>ARMENIA<br>AT SEA<br>ATLANTIC OCEAN<br>AZERBAIJAN<br>AZORES<br>BAHAMAS<br>BAHRAIN<br>BANGLADESH<br>BELARUS<br>LATVIA<br>LEBANON<br>LEEWARD<br>LIBERIA<br>LIBYA<br>LITHUANIA<br>MACAO<br>MALAYSIA<br>MARIANA<br>MAURITIUS<br>MOROCCO<br>MOZAMBIQUE<br>NETHERLAND<br>NETHERLANDS<br>NICARAGUA<br>NIGERIA<br>NORTH KOREA<br>OMAN<br>PAKISTAN<br>PANAMA |  |  |
|                                     | FRANCE<br>GERMANY<br>GREECE<br>HOLLAND<br>ITALY<br>JAPAN<br>MONACO<br>NETHERLANDS<br>NEW ZEALAND   |  |  |
|                                     | NORWAY<br>PORTUGAL<br>RUSSIA<br>SPAIN<br>SWEDEN<br>SWITZERLAND<br>TENERIFE<br>UK<br>USA  |  |  |
|                                     | GEORGIA<br>HONG KONG<br>HUNGARY<br>ISRAEL<br>MADERIA<br>MALTA<br>MÉXICO<br>PERU<br>POLAND  |  |  |
|                                     | ROMANIA<br>ROUMANIA<br>SLOVAKIA<br>SOUTH AFRICA<br>SOUTH KOREA<br>UKRAINE<br>VENEZUELA<br>YUGOSLAVIA   |  |  |
|                                     | GUATEMALA<br>GUYANA<br>INDIA<br>INDONESIA<br>IRAN<br>IRAQ<br>ISTANBUL<br>IVORY COAST<br>JAMAICA<br>JAVA<br>JORDAN<br>KARACHI<br>KAZAKHSTAN<br>KENYA<br>KOREA<br>KUALA LUMPUR<br>KUWAIT<br>TAIWAN<br>TANZANIA<br>THAILAND<br>TRINIDAD<br>TRINIDAD & TOBAGO<br>TUNISIA<br>TURKEY<br>UAE<br>UGANDA<br>UNITED ARAB EMIRA<br>URUGUAY<br>UZBEKISTAN<br>VIETNAM<br>VIRGIN ISLANDS<br>WEST AFRICA<br>WEST INDIES<br>YEMEN<br>ZAMBIA<br>ZIMBABWE                |  |  |

Taula 2: Taula dels estats classificats segons el seu grau de desenvolupament.



### 3.3.3. Depuració de les variables que indiquen les característiques de les substàncies involucrades en l'accident

- **MN** és categòrica i indica els noms de les substàncies que han intervingut en l'accident. Conté 100 missings, és a dir, es disposa dels noms de les substàncies que han participat en 5.076 accidents.
- **ESTAT\_FÍSIC** és categòrica i classifica el tipus de substància segons el seu estat físic en el moment de l'accident. Inicialment, té 11 categories i 973 missings. Es coneix l'estat físic de les substàncies que han participat en 4.201 accidents. Parlant amb l'enginyer responsable del projecte s'ha decidit recodificar aquesta variable en 5 categories no excloents per tal de fer disminuir el nombre de categories utilitzant una classificació més global. Aquestes categories són:
  - SÒLID codificat amb el nombre 1. S'han inclòs les categories: SÒLID, SLURRY i combinacions d'aquests estats físics.
  - LÍQUID codificat amb el nombre 2. S'han inclòs les categories: LÍQUID, SOLUTION, SUPERHEATL i combinacions d'aquests estats físics.
  - GAS codificat amb el nombre 3. S'han inclòs les categories: GAS, RLGAS, LGAS i combinacions d'aquests estats físics.
  - PLGAS codificat amb el nombre 4.
  - DUST codificat amb el nombre 5.

Com que aquestes categories no són excloents, s'ha decidit per no perdre informació, construir 5 variables indicadores no excloents. En el cas que hi hagi participat més d'una substància en el mateix accident es coneixeran els estats físics a partir d'aquestes variables indicadores i a partir de la variable que indica el nombre de substàncies que han intervingut en l'accident, **NSUS**. Per exemple, si en un accident han participat 3 substàncies i totes en estat líquid, la variable indicadora LÍQUID tindrà un 1 i la resta de variables indicadores un zero per

aquest accident, i a partir de la variable **NSUS** es coneixerà que hi ha participat en l'accident 3 substàncies i aquestes 3 en estat físic líquid. Si en un accident hi han intervingut dues substàncies, una en estat físic líquid i una altra en estat físic sòlid, apareixerà un 1 en la variable indicadora **LÍQUID** i un 1 en **SÒLID**, i a la resta de variables indicadores un zero i en la variable **NSUS** apareixerà un 2. També es pot donar el cas que hagin participat dues substàncies en un mateix accident i que no es coneix l'estat físic d'una d'elles, en aquest cas en la variable **NSUS** es trobarà un 2 i només apareixerà un 1 en una de les variables indicadores que indiquen l'estat físic.

Per construir aquestes variables indicadores, com que no són excloents, no es pot utilitzar la instrucció Manip/Code del programa MINITAB, així doncs, s'han hagut de fer-les manualment en el programa EXCEL i posteriorment, tenint en compte el codi identificatiu de cada accident, s'han copiat les noves variables a MINITAB.

Com amb el nom de la substància a vegades es pot deduir el seu estat físic, s'ha facilitat una llista a l'enginyer responsable del projecte amb el nom de totes les substàncies on no es coneixia l'estat físic per veure si es podia deduir segons criteris de tipus químic. D'aquesta manera s'ha reduït el nombre de missings en 374, ara aquesta variable només en té 599. Però, només s'ha pogut reduir el nombre de missing de les categories **SÒLID**, **LÍQUID** i **GAS**, sense poder assolir un nivell més alt de detall.

La variable **MH** de la base de dades original a partir d'ara s'anomena **HAZARD**.

- **HAZARD** registra les categories de perill de les substàncies que han intervingut en l'accident. A partir de les categories d'aquesta variable, s'han creat les 8 variables indicadores no excloents que es mostren a continuació:

- **HZTO** indica si la substància és tòxica. Hi ha 1.521 accidents on hi ha intervingut una o varies substàncies tòxiques.
- **HZFI** indica si la substància és inflamable. Hi ha 3.269 accidents on hi ha intervingut una o varies substàncies inflamables.
- **HZEX** indica si la substància és explosiva. Hi ha 444 accidents on hi ha intervingut una o varies substàncies explosives.
- **HZCO** indica si la substància és corrosiva. Hi ha 599 accidents on hi ha intervingut una o varies substàncies corrosives.
- **HZRA** indica si la substància és radioactiva. Hi ha 5 accidents on hi ha intervingut una o varies substàncies radioactives.
- **HZCD** indica si la substància és refrigerada. Hi ha 47 accidents on hi ha intervingut una o varies substàncies refrigerades.
- **HZAS** indica si la substància és asfixiant. Hi ha 32 accidents on hi ha intervingut una o varies substàncies asfixiants.
- **HZOX** indica si la substància és oxidant. Hi ha 307 accidents on hi ha intervingut una o varies substàncies oxidants.

Pot donar-se el cas de que la substància pertanyi a més d'una de les variables mencionades, ja que pot tenir més d'un risc, per exemple ser tòxica i inflamable. Aleshores apareixerà un 1 en les variables **HZTO** i **HZFI**. Aquestes variables no tenen dades mancants.

Cal recordar que la base de dades MHIDAS no inclou específicament substàncies radioactives, ja que hi ha bases de dades especialitzades en accidents produïts amb substàncies d'aquest tipus. En la base de dades adaptada només hi ha 4 accidents on ha participat la substància *URANIUM HEXAFLOURIDE* i un accident on ha participat la substància *NUCLEAR WASTE*, que s'han recollit perquè són tòxiques i a més a més són radioactives. Si haguessin estat únicament radioactives no s'haurien recollit. Per tant, el risc radioactiu no és representatiu de les substàncies químiques perilloses que formen la base de dades.

- **MC** és el codi utilitzat per identificar la substància o la família a la que pertany, és per això que es pot trobar el mateix codi per diferents substàncies. Quan en un accident ha participat més d'una substància, els codis d'aquestes aniran separats per una barra. Per exemple: 9999/1114/9999 ó /1267. Aquest últim exemple indica que en aquest accident han intervingut dues substàncies, però que es desconeix el codi de la primera. El codi 9999 vol dir que la substància pertany a la categoria ALTRES. Aquesta variable categòrica té 42 missings dels quals en 4 no es coneix el codi de la primera substància, per tant, es disposa dels codis de 5.134 accidents en els que es coneixen totes les substàncies que han participat.
  
- **NSUS** és discreta i indica el nombre de substàncies que han intervingut en l'accident. El rang de valors és de 1 a 6. La majoria són accidents on hi ha intervingut una única substància, 4.646 accidents. Després n'hi ha 371 accidents en els que hi ha intervingut 2 substàncies, 102 en els que hi ha intervingut 3, 46 en els que hi ha intervingut 4, 8 en els que hi ha intervingut 5 i 3 accidents en els que hi ha participat 6 substàncies. Aquesta variable no té dades mancants.

### 3.3.4. Depuració de les variables que descriuen el tipus d'accident

A MHIDAS es pot trobar la variable categòrica **IT** que indica mitjançant dos codis el tipus d'accident de forma seqüencial: DESCRIPCIÓ PRINCIPAL 1 – DESCRIPCIÓ SECUNDÀRIA 1, DESCRIPCIÓ PRINCIPAL 2 – DESCRIPCIÓ SECUNDÀRIA 2. On la descripció principal indica el tipus d'accident de forma general i la secundària de forma més específica. Per reduir el nombre de categories i desglossar la descripció principal de la secundària, s'ha dividit aquesta variable en dues:

- **TIPUSACC** descriu el tipus d'accident, en el pas del temps, mitjançant les descripcions principals. Conté 205 missings. Es coneixen les descripcions principals de 4.971 accidents. Té aquestes quatre categories no excloents: FIRE, EXPLODE, RELEASE i GASCLD.

Aquesta variable pot tenir una d'aquestes quatre categories o una combinació de dues d'elles com a màxim. Per exemple: EXPLODE-FIRE. A partir de les quatre categories, s'han construït 4 variables indicadores amb els noms **TACCEX**, **TACCFI**, **TACCGAC** i **TACCREL**. Per construir aquestes variables indicadores, que no són excloents, no es pot utilitzar la instrucció Manip/Code del programa MINITAB, i s'han hagut de fer manualment amb el programa EXCEL. Posteriorment, s'han copiat les noves variables a MINITAB tenint en compte el codi identificatiu de cada accident.

- **IT2** descriu el tipus d'accident, si existeix aquesta informació (sinó posa el mateix que a **TIPUSACC**), en el pas del temps, mitjançant els codis secundaris. Conté els mateixos missings que **TIPUSACC**, 205, per tant, es coneixen els descriptors secundaris de 4.971 accidents. Té les següents categories o una combinació de dues d'elles com a màxim: BLEVE, CONTREL, DENSEXP, DENSGSCL, DUSTEXP, EXPLODE, FIRE, GASCLD, INSTREL, MISTEXP, NEUGSCL, PHYSEXP, POOL, RELEASE i TORCH. Aquesta variable no s'utilitzarà pel seu gran nombre de categories.
- **IS** és categòrica i indica la font d'ignició de l'accident. Pot tenir fins a dos codis per descriure-la. Té 4.111 missings, és a dir, es coneix la font d'ignició de 1.065 accidents. Té les 17 categories següents: AUTOIGNITE, ELECTRIC-, DOMESTICS, ELECTRIC-INSTRUMENT, ELECTRIC-MOTorgen, ELECTRIC-STATICS, ELECTRIC-VEHICLE, FLAMES-BOILER, FLAMES-FLARE, FLAMES-MATCH, FLAMES-STOVE, FRICSPARK-MACHINERY, FRICSPARK-SPARKTOOL, HOTSURFACE-FRICSURF, HOTSURFACE- INCAND, HOTSURFACE- STEAMPIPE, HOTSURFACE-VEXHAUSTS i NONIGNITE. Aquesta variable no es farà servir en l'estudi pel gran nombre de dades mancants i la gran quantitat de categories que presenta.

- **IG** és discreta i conté una estimació a la baixa o a l'alça del temps, en segons, transcorregut entre el començament de la fuga i la ignició en aquells accidents per explosió o incendi. Aquesta variable té molts zeros a la base de dades d'accident que són missings en la base de dades de substàncies. Per tal de fer coincidir els missings de les dues bases de dades, s'ha fet una consulta d'actualització mitjançant el programa de gestió de bases de dades ACCESS. D'aquesta manera ara els zeros que conté la base de dades adaptada són la veritable estimació del temps entre la fuga i la ignició. El rang de valors és de 0 a 18.000 segons. Té 4.929 missings, la qual cosa vol dir que només es coneix el temps entre el començament de la fuga i la ignició de 247 accidents. Aquesta variable no s'utilitzarà en l'estudi degut a l'elevat nombre de missings que presenta.

### 3.3.5. Depuració de les variables que expliquen les causes de l'accident

A MHIDAS la variable **OG** descriu el tipus d'activitat que es duia a terme en el moment de l'accident i el lloc on es va originar l'incendi mitjançant dos codis. Per reduir el nombre de categories, s'ha dividit aquesta variable en dues:

- **TIPUSACT** que recull el tipus d'activitat. Té 187 missings, és a dir, es coneixen els tipus d'activitats de 4.989 accidents i es poden trobar les següents 6 categories: DOM/COM, PROCESS, STORAGE, TRANSFER, WAREHOUSE i WASTE.

Com que les categories WAREHOUSE i WASTE estan molt relacionades, ja que ambdues es refereixen a tasques auxiliars de la indústria, s'ha decidit d'acord amb l'enginyer responsable del projecte que és convenient agrupar-les en la categoria ALTRES, i d'aquesta manera es tindrà una categoria menys.

A partir de les categories d'aquesta variable s'han creat cinc variables indicadores amb els noms **TACTDO/C**, **TACTPRO**, **TACTSTO**, **TACTTRANS** i **TACTAL**, encara que només s'utilitzaran 4 per evitar problemes de colinealitat.

- **OG2** recull l'origen específic. Té 1.626 missings, per tant, es coneixen els orígens específics de 3.550 accidents. Té les 22 categories següents: **ASVESSEL**, **BARGE**, **COMMTANK**, **FIREDEQUIP**, **HEATXCHANG**, **HOSE**, **MACDRIVE**, **PACKAGE**, **PIPELINE**, **PIPEWORK**, **PSVESSEL**, **PUMP**, **PVESSEL**, **RAILTANKER**, **REACTOR**, **ROADTANKER**, **SHIP**, **SIZECHANGE**, **SOLIDMOVE**, **SOLIDSTORE**, **SUBSTATION** i **TANKCONTNR**. Aquesta variable no s'utilitzarà pel gran nombre de categories que presenta.

La variable **GC** de la base de dades original a partir d'ara l'anomenarem **CAUSAG**.

- **CAUSAG** és categòrica i indica la causa general de l'accident. Té 2.566 missings, per tant, es coneix la causa general de 2.610 accidents. Aquesta variable no excloent té les següents 8 categories: **EXTERNAL**, **HUMAN**, **IMPACT**, **INSTRUMENT**, **MECHANICAL**, **PROCOND**, **SERVICE** i **VREACTION**.

Tot i que semblen poques categories, com que hi ha la possibilitat de què hi intervingui més d'una causa, es poden trobar combinacions. Per exemple: **VREACTION** i **HUMAN**. La categoria **EXTERNAL** té una variant que és **EXTERNAL'** que correspon a la mateixa categoria, **EXTERNAL**, ja que es va produir un error en passar les dades de la base original a l'adaptada. S'han trobat 13 casos amb aquest error que tenen en comú la categoria **SABOTAGE** a la variable **SC** (causa específica). S'ha solucionat aquest problema codificant **EXTERNAL'** com a **EXTERNAL**.

A partir de les categories d'aquesta variable s'han construït 8 variables indicadores no excloents amb els noms: **CGEXT**, **CGHUM**, **CGIMP**, **CGMEC**, **CGVREA**, **CGINST**, **CGPROC** i **CGSERV**. Per construir aquestes variables indicadores, que no són excloents, no es pot utilitzar la instrucció Manip/Code del programa MINITAB, i s'han hagut de fer manualment amb el programa EXCEL. Posteriorment, s'han copiat les noves variables a MINITAB, tenint en compte el codi identificatiu de cada accident. Tot i que les categories MECHANICAL i INSTRUMENT segons l'enginyer responsable del projecte estan molt relacionades s'ha decidit estudiar-les per separat ja que no sempre quan la causa de l'accident ha estat mecànica també hi ha hagut una fallida d'instrumentació.

- ❑ **SC** és categòrica i descriu les causes específiques de l'accident. Té 3.035 missings, és a dir, es coneixen les causes específiques de 2.141 accidents. Aquesta variable té les 54 categories següents: OVERHEAT, OVERLOAD, OVERPRES, CORRODE, WELDFAIL, FATIGE, METALLURG, INCOMPAT, RELIEFVALV, HOSE, BRITTLE, HVYOBJECT, CRANE, EXCAVEQUIP, VEHICLE, ROADACC, RAILACC, SHIP/SHIP, SHIP/LAND, GENERALOP, OVERFILL, DRAINACC, ACCVENT, MAINTAIN, MANAGEMENT, COMMUNICAT, PROCEDURES, DESIGN, INSTALL, CONNECTING, CONSTRUCT, CONTROL, INDICATOR, ALARM, TRIP, ELECTRIC, WATER, GAS, GLANDSEAL, COMPAIR, STEAM, INTNLFIRE, CONEXP, RUNAWAY, EXTNLFIRE, EXTNLEXP, EARTHQUAKE, GROUND, FLOODS, LIGHTNING, HIGHWINDS, TEMPRTURE, SABOTAGE i VALVE. Aquesta variable no s'utilitzarà en l'estudi ja que presenta un gran nombre de categories i de missings.



### 3.3.6. Depuració de la variable que mesura la quantitat de substància involucrada en l'accident

A MHIDAS es té la variable **QY**, però aquesta no es pot considerar continua, ja que és de text, i presenta els mateixos problemes que **KILLED**, **INJURED**, **EVACUATED** i **DAMAGE**. Es pot trobar, per exemple, que la quantitat de substància intervinguda és  $>0,001$  ó  $<50$ . Per passar-la a numèrica s'ha prescindit dels símbols de desigualtat. Per aquest motiu es té a la base de dades adaptada la variable **QSUS** que s'ha creat a partir de la variable **QY**.

**QSUS** conté la suma de la quantitat de substància o substàncies involucrada en l'accident. Aquesta variable té 4.023 zeros que són missings en la base de dades original i 3 que ho són en realitat (amb AN 2742, 4405 i 4360), per tant, per solucionar aquesta errada s'ha necessitat la base de dades de substàncies per comparar-la amb la base de dades d'accidents, això s'ha fet mitjançant una consulta d'actualització amb el programa de gestió de bases de dades ACCESS i el que s'ha fet ha sigut codificar amb "\*" els registres d'aquesta columna que vertaderament eren dades mancants. Però parlant amb l'enginyer responsable del projecte s'ha arribat a la conclusió de que és impossible que en un accident intervingui una quantitat de substància igual a zero. Per això s'ha decidit codificar com a missing els tres zeros abans mencionats. A més a més s'ha construït una nova variable per tal de conservar **QY** tal com estava originalment, codificada amb 1, 2 i 3, tal i com s'ha fet amb les variables resposta.

- **QSUS** estima la quantitat de material involucrat en l'accident en tones. Té 4.026 missing codificats mitjançant un "\*", per tant, es coneix la quantitat de substància que ha intervingut en 1.150 accidents. El rang de valors és de 0,001 a 99.999<sup>1</sup>.

---

<sup>1</sup> Aquest 99999 és el màxim de la base de dades MHIDAS, per tant, aquest valor indica com a mínim aquesta quantitat de substància involucrada.

### 3.3.7. Depuració de les variables que expliquen l'accident

Les variables que expliquen l'accident són de text i no es poden categoritzar, ja que hi ha una explicació per cada accident, per tant, no s'utilitzaran. Aquestes variables són:

- ❑ **AB** fa un breu resum de l'accident. Només té un missing (per tant, es disposa dels resums de 5.175 accidents) que correspon a l'accident amb AN 5637.
- ❑ **KW** indica a través d'unes paraules clau acompanyades d'un text si existeix més informació sobre alguns aspectes addicionals. Per exemple: AFTER. Té 1.164 missings, és a dir, es disposa de les paraules clau de 4.012 accidents.

### 3.3.8. Depuració de les variables complementàries

- ❑ **AN** és discreta i indica el codi de l'accident. Hauria de tenir un codi per cada accident diferent, però hi ha accidents aparentment diferents que tenen el mateix codi. Per solucionar aquest problema s'ha parlat amb l'enginyer responsable del projecte per tal d'adoptar la solució més adient en cada cas i fer que només hi hagi un codi per accident. A continuació es descriu la solució adoptada per a cada codi repetit:

El codi AN 1023 està repetit dues vegades, però només es refereix a un accident. El dia, el mes, l'any, la regió, l'estat, el tipus d'accident, l'activitat i l'origen, el resum de l'accident i les paraules clau són iguals en tots dos registres. La resta de variables no es coneixen. Les úniques diferències són la localitat (en un es coneix i en l'altre no), el nombre de persones evacuades, en un posa 40 i en l'altre 1 i les variables que són típiques de cada substància, com el nom, el codi, etc. L'error, per

tant, s'ha produït perquè un accident tenia especificada la localitat "Jersey City" i l'altre no. Això és degut a que el programa VISUAL BASIC, que s'ha utilitzat per classificar les substàncies per accidents, els ha considerat accidents diferents.

Per tal de conèixer el valor correcte de persones evacuades s'ha consultat la base de dades original (de substàncies) i s'ha vist que de les tres substàncies que havien intervingut en l'accident dues d'elles tenien el mateix nombre d'evacuats, és a dir, 40. A més a més, la descripció de les tres substàncies indicava que havien sigut evacuats 40 habitatges. Per tant, s'ha decidit agrupar en un mateix accident aquests dos registres, considerant que el nombre d'evacuats és 40 i la localitat "Jersey City". A més a més, s'afegeix el nom, l'estat físic i el codi de la tercera substància i es canvia el nombre de substàncies que han intervingut en l'accident.

El codi AN 1639 està repetit dues vegades, però només es refereix a un accident. El mes, l'any, la localitat, l'estat, el tipus, l'activitat i l'origen, la causa general, la causa específica, el nombre de morts, el nombre de ferits, el resum i les paraules clau de l'accident són iguals. Les diferències estan en què en un es coneix la regió, "Maine", i en l'altre no. L'error, per tant, s'ha produït perquè un accident tenia especificada la regió i l'altre no. Per tant, s'agrupen aquests dos registres en un, deixant com a regió, la coneguda. A més a més, s'afegeix el nom, l'estat físic i el codi de la segona substància, es canvia el nombre de substàncies que han intervingut en l'accident, i s'afegeix que la darrera substància és inflamable.

El codi AN 179 està repetit dues vegades, però només es refereix a un accident. El dia, el mes, l'any, la regió, l'estat, el tipus, el nombre de morts, el nombre de ferits i el nombre d'evacuats de l'accident són iguals en tots dos registres. Les diferències estan en què en un s'especifica com a localitat "Big Spring" i en l'altre "Big Springs" (la mateixa però afegint al final una -s) i l'estimació econòmica del material en un és 25 i en l'altre 6,8 milions de dòlars. També es diferencien en l'activitat i l'origen de l'accident, en un s'especifica i en l'altre no, el mateix passa amb la causa específica i la general. En referència a les paraules clau i el resum de l'accident, una descripció engloba la descripció de l'altre. El contingut de les variables que són típiques de cada substància tal com el nom i el codi, també són diferents. L'error, per tant, s'ha produït perquè s'han considerat localitats diferents.

Per tant, s'agrupen aquests dos registres en un de manera que la localitat és "Big Spring". En quan a l'activitat, l'origen, la causa general i l'específica de l'accident, es deixen les conegudes. Pel que fa al resum i a les paraules clau de l'accident es deixa la descripció més detallada. En l'estimació econòmica del dany material es posa 15,9 milions de dòlars US, que és la mitjana, ja que es desconeix el valor correcte. A més a més, s'afegeix el nom, l'estat físic i el codi de la segona substància, es canvia el nombre de substàncies que han intervingut en l'accident i s'afegeix que aquesta darrera substància és inflamable.

El codi AN 1941 està repetit dues vegades, i com que es diferencien en el dia, el mes, el resum, les paraules clau, la causa general, la causa específica, la font d'ignició, el nombre de ferits de l'accident, a més a més del contingut de les variables que són típiques de cada substància, es considera que són dos accidents diferents, tot i que pertanyen al mateix any, localitat, regió i estat. Per tal de diferenciar aquests dos accidents a un se li deixa aquest codi, 1941, i a l'altre se li posa el codi 10000, ja que com només hi ha codis de 4 xifres a la base de dades, el codi 10000 no existia amb anterioritat i serà propi d'aquest accident.

El codi AN 2152 està repetit dues vegades, però només es refereix a un accident. El dia, el mes, l'any, la regió, la densitat de població, l'estat, l'activitat i l'origen, la causa general, la causa específica, el nombre de quantitat de substància intervinguda, el nombre de persones evacuades, el nombre de morts, el nombre de ferits, les paraules clau i el resum de l'accident són iguals en tots dos registres. Les diferències estan en la variable localitat, en un és "Quebec" i en l'altre "Nitro" i en les variables que són típiques de cada substància tal com el nom i el codi. L'error, per tant, s'ha produït perquè hi ha localitats diferents. Per tant, s'agrupen aquests dos registres en un de manera que es deixa com a localitat "Quebec", ja que no es té constància bibliogràfica de que existeixi "Nitro" com a localitat. També s'afegeix el nom, l'estat físic i el codi de la segona substància i es canvia el nombre de substàncies que han intervingut en l'accident, tot afegint que la darrera substància és corrosiva i oxidant.

El codi AN 2322 està repetit dues vegades, però només es refereix a un accident. El dia, el mes, l'any, la localitat, l'estat, la densitat de població, el tipus, l'activitat i l'origen, la causa general, l'estimació econòmica del dany material, el nombre de morts, el nombre de ferits, el nombre d'evacuats, les paraules clau i el resum de l'accident són iguals en tots dos registres. Les diferències estan en la variable regió, ja que en un és "Texas" i en l'altre "Taxas", i en les variables que són típiques de cada substància tal com el nom i el codi. L'error, per tant, s'ha produït perquè hi ha regions diferents. Per tant, s'agrupen aquests dos registres en un de manera que la regió és "Texas". A més a més, s'afegeix el nom, l'estat físic i el codi de la tercera substància i es canvia el nombre de substàncies que han intervingut en l'accident.

El codi AN 2483 està repetit dues vegades, però fa referència al mateix accident. El dia, el mes, l'any, la localitat, l'estat, la densitat de població, la causa general, el nombre de morts, el nombre de ferits, el nombre d'evacuats, les paraules clau i el resum de l'accident són iguals en tots dos registres. Les diferències estan en la variable regió, ja que en un posa "Liguria" i en l'altre és missing, en el tipus d'accident, ja que en un hi ha EXPLODE-FIRE i en l'altre FIRE-EXPLODE. L'activitat és la mateixa, però l'origen no, ja que en un hi ha ASVESSEL i en l'altre PSVESSEL. Pel que fa a la causa específica una engloba l'altre, la causa general i les variables que són típiques de cada substància tal com el nom i el codi, també són diferents. L'error, per tant, s'ha produït perquè un tenia com a regió "Liguria" i l'altre no en tenia cap. Per tant, s'agrupen aquests dos registres en un de manera que es deixa com a regió la que es coneix, com a tipus d'accident EXPLODE-FIRE, ja que segons el resum de l'accident primer és va produir l'explosió i després l'incendi. En l'origen es posa la combinació dels dos, ASVESSEL-PSVESSEL, i com a causa específica la descripció més detallada. A més a més, s'afegeix el nom, l'estat físic i el codi de la segona substància, i es canvia el nombre de substàncies que han intervingut en l'accident.

El codi AN 2901 està repetit dues vegades, però només es refereix a un accident. El dia, el mes, la localitat, la regió, l'estat, el tipus, la causa general, la causa específica, l'estimació econòmica del dany material, el nombre de morts, el nombre de ferits, el nombre d'evacuats, les paraules clau i el resum de l'accident són iguals

en tots dos registres. Les diferències són l'any, en un és 77 i en l'altre 87, el tipus d'activitat, en un es desconeixia l'activitat i en l'altre posava STORAGE, l'origen, en un es desconeixia l'origen i en l'altre posava ASVESSEL, i les variables que són típiques de cada substància tal com el nom i el codi. L'error, per tant, s'ha produït perquè hi ha dos anys diferents pel mateix accident, però com que en aquest accident hi ha participat tres substàncies i en la base de dades de substàncies dues tenien l'any 87 i l'altre 77, es deixa l'any 87. Per tant, s'agrupen aquests dos registres en un de manera que es deixa com any el 87, i com a activitat i origen els que es coneixen. A més a més, s'afegeix el nom, l'estat físic i el codi de la tercera substància i es canvia el nombre de substàncies que han intervingut en l'accident.

El codi AN 398 està repetit dues vegades, però es refereix al mateix accident. El dia, el mes, la densitat de població, la localitat, l'estat, l'activitat, l'estimació econòmica del dany material, el nombre de morts, el nombre de ferits, el nombre d'evacuats, les paraules clau i el resum de l'accident són iguals en tots dos registres. Les diferències són l'any, en un és 78 i en l'altre 79, el tipus d'accident, en un només s'especifica FIRE i en l'altre FIRE-EXPLODE, l'origen, en un posava COMMTANK i en l'altre posava SOLIDSTORE, i les variables que són típiques de cada substància tal com el nom i el codi. L'error, per tant, s'ha produït perquè hi ha dos anys diferents pel mateix accident, però com pertanyen a la mateixa dècada i es farà servir aquesta variables en comptes de l'any, és indiferent l'any que es deixi. Per tant, s'agrupen aquests dos registres en un de manera que es deixa com any el 78, per exemple, en el tipus d'accident es deixarà FIRE-EXPLODE, en comptes de FIRE, perquè en el resum s'especifica que primer es va produir un incendi que va provocar una explosió posterior i en l'origen es deixa la combinació dels orígens que es coneixen, és a dir, COMMTANK-SOLIDSTORE. A més a més, s'afegeix el nom, l'estat físic i el codi de la tercera substància i es canvia el nombre de substàncies que han intervingut en l'accident.

El codi AN 6232 està repetit dues vegades, però fa referència a un únic accident. El dia, el mes, l'any, la localitat, l'estat, el tipus, l'activitat i l'origen, el nombre de ferits, les paraules clau i el resum de l'accident són iguals en els dos registres. Les diferències estan en la regió, ja que en un és "Java" i en l'altre és missing, en la

causa general i l'específica, ja que en un es coneix i en l'altre no. També es diferencien en les variables que són típiques de cada substància tal com el nom i el codi. L'error, per tant, s'ha produït perquè un tenia com a regió "Java" i l'altre no. Per tant, s'agrupen aquests dos registres en un de manera que es deixa com a regió, causa general i causa específica les mencionades. A més a més, s'afegeix el nom, l'estat físic i el codi de la segona substància i es canvia el nombre de substàncies que han intervingut en l'accident.

El codi AN 675 està repetit dues vegades, però no fa referència al mateix accident, ja que van succeir en dies diferents, encara que en el mateix mes, any i localitat. L'activitat i l'origen, tipus d'accident, l'estimació econòmica del dany material, el nombre de ferits i el nombre d'evacuats també són iguals. Les diferències estan en la causa general i l'específica, el nombre de morts, les paraules clau i el resum de l'accident. Per això, es considera que són accidents diferents. Per tal de diferenciar aquests dos accidents a un se li deixa aquest codi i a l'altre se li posa el codi 10001, ja que només hi ha codis de 4 xifres a excepció del que s'ha modificat abans (10000).

Els altres codis repetits no mencionats en aquest apartat tenen com a activitat TRANSPORT i com que aquests no interessen per l'estudi no s'han considerat. La variable AN no té missings. Després de realitzar les modificacions especificades s'ha passat a disposar de 5.176 a 5.167 accidents industrials en els que han participat substàncies químiques perilloses.

Les 2 variables complementaries que s'expliquen a continuació no es faran servir en l'estudi perquè no són importants per explicar les variables resposta.

- **CR** indica la font d'informació de l'accident. Hi ha 35 missings, és a dir, que es disposa de la font d'informació de 5.141 accidents. Aquesta és una variable qualitativa amb aquestes 9 categories:

- AEA
- DURRAN
- HSE: Health and Safety Executive
- ICHIME
- KAAT
- OECD: Organització de Cooperació i Desenvolupament Econòmic.
- PAHO
- SRD: Safety and Reliability Directorate
- WORKSA

No s'ha pogut trobar el significat de totes aquestes sigles.

- **RA** és de text i indica el nombre d'articles o textos disponibles sobre l'accident que es poden consultar. Com a valors s'hi pot trobar per exemple: >1, >60 i 9. Té 18 missings. Es disposa de textos sobre 5.158 accidents.



### 3.4. Resum del més important

En aquest projecte s'ha limitat l'àmbit d'estudi als accidents produïts durant el desenvolupament de tasques en instal·lacions fixes, i per tant s'han eliminat de la base de dades aquells accidents que es van produir durant el transport, ja sigui per terra o per mar. Això ha fet que la base de dades adaptada quedés reduïda a 5.176 registres.

També s'han transformat les variables de text **DAMAGE**, **KILLED**, **INJURED**, **EVACUATED** i **QY** a numèriques. També s'han corregit molts errors de codificació dels missings, ja que diverses variables contenien molts zeros que en realitat eren dades mancats. Aquest problema s'ha solucionat mitjançant consultes d'actualització amb el programa de gestió de base de dades ACCESS. Les variables resposta **KILLED**, **INJURED**, **EVACUATED** i **DAMAGE**, el temps transcorregut entre la fuga i la ignició, **IG**, i la quantitat de substància involucrada, **QY**, presentaven aquest problema.

S'han creat diverses variables indicadores que s'utilitzaran més endavant quan es construeixin els models per a les variables resposta. A més a més, s'han creat variables per actualitzar l'estimació del dany material, i d'aquesta manera poder comparar accidents d'anys diferents. S'ha decidit que s'utilitzarà la variable **DAMAGE-MAR**, que conté l'actualització econòmica del dany material mitjançant l'índex d'actualització del cost Marshall & Stevens, per estudiar el dany econòmic, ja que és la variable econòmica que presenta més dades actualitzades. A partir d'ara aquesta variable s'anomena **DAMAGE**.

Per tal de tenir una variable geogràfica amb poques categories, s'ha creat la variable **MÓN**, que classifica els estats segons el seu grau de desenvolupament econòmic. Aquesta variable té 3 categories.

S'han corregit tots els registres que tenien **AN**, codi d'accident, repetits, consultant i comparant la base de dades d'accidents amb la de substàncies. En alguns casos, quan era evident que hi havia un error i que els dos registres pertanyien al mateix accident s'han tornat a agrupar en un únic registre. Quan no hi havia cap motiu lògic per considerar que els dos registres feien referència al mateix accident, no s'han agrupat, i el que s'ha fet per a diferenciar-los ha estat canviar el codi d'un accident. Amb aquesta operació s'ha quedat reduïda la base de dades adaptada a 5.167 accidents.

Cal recordar que hi ha variables que contenen molts missings i d'altres que tenen moltes categories. Per tal de reduir el nombre de categories de les variables, d'acord amb l'enginyer químic responsable del projecte, s'han recategoritzat variables utilitzant una classificació més general. Aquest és el cas de la variable **MT**, estat físic de la substància en el moment de l'accident, inicialment tenia 11 categories i ara, amb la nova classificació, en té 5, i s'anomena **ESTAT\_FÍSIC**.

Per tal de seleccionar les variables més importants per construir els models de regressió s'ha consultat amb l'expert i s'ha arribat a la conclusió que les variables explicatives més importants que es disposen per explicar les variables resposta són: L'estat físic de la substància, **ESTAT\_FÍSIC**, les variables que indiquen el risc de la substància o substàncies com són **HZTO**, **HZFI**, **HZEX**, **HZCO**, **HZRA**, **HZCD**, **HZAS** i **HZOX**, el tipus d'activitat, **TIPUSACT**, la causa general de l'accident, **CAUSAG**, el tipus d'accident **TIPUSACC**, l'antiguitat, **ANTIGUITAT**, la dècada, **DÈCADA**, la quantitat de substància involucrada en l'accident, **QSUS**, i el nombre de substàncies que han participat en l'accident, **NSUS**.

La variable **MÓN** que s'ha creat pot ser que no s'introdueixi en el model de regressió si aquesta fa que una altre variable més important que ella no entri en el model.

Les altres variables queden, a priori, descartades. Com per exemple les variables **IS**, font d'ignició, i **IG**, temps d'ignició, ja que aquestes variables presenten moltes dades mancants.

Pel que fa al cas de les variables qualitatives, amb moltes categories, com per exemple: **IT2** i **OG2**, han sigut descartades pel gran nombre de categories que tenen, ja que es tracta de categories molt específiques i la seva classificació general ja es té en **TIPUSACC** i **TIPUSACT** respectivament. Les variables **IG** i **IS** han sigut descartades també, pel gran nombre de missings que tenen i per últim, també s'han descartat les variables de text perquè no són adients per la construcció de models de regressió.

A continuació es mostra una taula resum amb la descripció de les variables de la base de dades adaptada.

| VARIABLES                                     | CATEGORIES ORIGINALS | CATEGORIES PROPOSADES | RANG         | ZEROS | MISSINGS | NO MISSINGS        |
|---|----------------------|-----------------------|--------------|-------|----------|--------------------|
| KILLED  | NUMÉRICA             | -                     | 0-2.000      | 1.037 | 3.024    | 2.143              |
| INJURED                                       | NUMÉRICA             | -                     | 0-9.999      | 779   | 2.374    | 2.793              |
| EVACUATED                                     | NUMÉRICA             | -                     | 0-200.000    | 276   | 4.055    | 1.112              |
| DAMAGEI (estimació econòmica no actualitzada) | NUMÉRICA             | -                     | 0-370        | 5     | 4.409    | 758                |
| DAMAGE-CEI (e. econòmica amb l'índex CEI)     | NUMÉRICA             | -                     | 0,002-404,91 | -     | 4.433    | 734                |
| DAMAGE (e. econòmica amb l'índex MAR.)        | NUMÉRICA             | -                     | 0,002-456,53 | -     | 4.417    | 750                |
| INDEX-CEI                                     | NUMÉRICA             | -                     | 0-389,5      | -     | 0        | 5.176              |
| INDEX-MAR                                     | NUMÉRICA             | -                     | 0-1.061,9    | -     | 0        | 5.176              |
| DIA   | NUMÉRICA             | -                     | 1-31         | -     | 306      | 4.861              |
| MES   | NUMÉRICA             | -                     | 1-12         | -     | 141      | 5.026              |
| ANY   | NUMÉRICA             | -                     | 0-98         | -     | 0        | 5.167              |
| ANTIGUITAT (0:abans, 1:després 1.970)         | INDICADORA           | -                     | -            | 499   | 0        | 4.668 <sup>2</sup> |
| DECADA  | -                    | 10                    | -            | -     | 0        | 5.167              |
| PD (densitat de població)                     | 3                    | -                     | -            | -     | 3.924    | 1.243              |
| LOCATION (localitat)                          | 2.796                | -                     | -            | -     | 192      | 4.975              |
| REGION (regió)                                | 515                  | -                     | -            | -     | 1.566    | 3.611              |
| COUNTRY (estat)                               | 134                  | -                     | -            | -     | 32       | 5.135              |
| MÓN (nivell desenvolupament)                  | -                    | 3                     | -            | -     | 0        | 5.167              |
| MN (nom de la substància)                     | 1.373*               | -                     | -            | -     | 100      | 5.067              |
| ESTAT FÍSIC                                   | 11                   | 5                     | -            | -     | 599      | 4.568              |
| HZTO (risc tòxic)                             | INDICADORA           | -                     | -            | 3.646 | 0        | 1.521 <sup>2</sup> |
| HZFI (risc inflamable)                        | INDICADORA           | -                     | -            | 1.898 | 0        | 3.269 <sup>2</sup> |
| HZEX (risc explosiu)                          | INDICADORA           | -                     | -            | 4.723 | 0        | 444 <sup>2</sup>   |
| HZCO (risc corrosiu)                          | INDICADORA           | -                     | -            | 4.568 | 0        | 599 <sup>2</sup>   |
| HZRA (risc radioactiu)                        | INDICADORA           | -                     | -            | 5.162 | 0        | 5 <sup>2</sup>     |
| HZCD (risc refrigerat)                        | INDICADORA           | -                     | -            | 5.120 | 0        | 47 <sup>2</sup>    |
| HZAS (risc asfixiant)                         | INDICADORA           | -                     | -            | 5.135 | 0        | 32 <sup>2</sup>    |
| HZOX (risc oxidant)                           | INDICADORA           | -                     | -            | 4.860 | 0        | 307 <sup>2</sup>   |
| MC (codi de la substància)                    | 689*                 | -                     | -            | -     | 38       | 5.129              |
| NSUS (nombre substàncies)                     | NUMÉRICA             | -                     | 1-6          | -     | 0        | 5.167              |
| TIPUSACC (tipus d'accident)                   | 4<br>19*             | -                     | -            | -     | 205      | 4.962              |
| IT2 (tipus d'accident secundari)              | 15<br>89*            | -                     | -            | -     | 205      | 4.962              |
| IS (font d'ignició)                           | 17                   | -                     | -            | -     | 4.103    | 1.064              |
| IG (temps d'ignició)                          | NUMÉRICA             | -                     | 0-18.000     | -     | 4.920    | 247                |
| TIPUSACT (tipus d'activitat)                  | 7                    | 5                     | -            | -     | 185      | 4.982              |
| OG2 (origen específic)                        | 22                   | -                     | -            | -     | 1.622    | 3.545              |
| CAUSAG (causa general de l'accident)          | 8<br>115*            | 7                     | -            | -     | 2.562    | 2.605              |
| SC (causa específica de l'accident)           | 54<br>425*           | -                     | -            | -     | 3.032    | 2.137              |
| QSUS (quantitat substància)                   | NUMÉRICA             | -                     | 0,001-99.999 | 0     | 4.026    | 1.141              |
| AB (resum accident)                           | TEXT                 | -                     | -            | -     | 1        | 5.166              |
| KW (text amb paraules clau)                   | TEXT                 | -                     | -            | -     | 1.164    | 4.003              |
| AN (codi accident)                            | NUMÉRICA             | -                     | 1-10.001     | -     | 0        | 5.167              |
| CR (font informació)                          | 9                    | -                     | -            | -     | 35       | 5.132              |
| RA (nombre articles)                          | TEXT                 | -                     | -            | -     | 17       | 5.150              |

/

Taula 3: Descripció de les variables de la base de dades adaptada.

<sup>2</sup> En les variables indicadores aquest nombre és el nombre d'uns.

\* Nombre de combinacions.

# CAPÍTOL 4: ANÀLISI EXPLORATÒRIA DE DADES

## 4.1. Introducció

En aquest capítol s'estudien les variables que s'utilitzaran per construir els models amb tècniques univariants i bivariants. Primerament, es descriuen les variables resposta **KILLED**, nombre de persones mortes, **INJURED**, nombre de ferits, **EVACUATED**, nombre d'evacuats i **DAMAGE**, estimació econòmica del dany material (actualitzada amb l'índex de cost Marshall & Stevens), mitjançant diagrames de punts i boxplots, tant per elles com per la seva transformació (logaritme neperià). Amb aquestes eines s'hi pot saber la variabilitat, el rang, la influència dels zeros (ja que aquestes variables, com la majoria, en tenen molts) i les dades mancants.

A continuació, es descriuen les variables explicatives mitjançant diagrames de Pareto i diagrames de Pastís per a les variables categòriques, i diagrames de punts i boxplots per la variable continua, per poder localitzar quines categories són les més importants, i poder fer una categorització millor que la proposada o que la que hi havia, o simplement comprovar que la categorització proposada ja és adequada.

Posteriorment, i amb la finalitat de veure les relacions existents entre les variables resposta i les variables explicatives, es fan plots per la variable continua i gràfics bivariants i boxplots múltiples per les variables categòriques. Amb aquests es veurà quines categories tenen un comportament similar o diferent, i es podrà observar quina categoria produeix més o menys variabilitat per cadascuna de les variables resposta, a més a més de veure quina fa augmentar o disminuir les variables resposta.

Cal destacar que els gràfics bivariants són l'eina més útil i rellevant per veure relacions entre dues variables però és una eina limitada. Els diagrames bivariants projecten totes les dades sobre un pla i és probable que en aquesta projecció no s'hi detectin totes les dependències multivariants existents.

La informació que s'extraurà de fer aquesta anàlisi exploratòria de dades serà de molta utilitat per explicar el comportament de les variables resposta i explicatives, i veure les relacions que existeixen entre elles. També es podran detectar dades atípiques, que tenen un comportament diferent a la resta d'accidents, és a dir, un valor massa alt o baix de la variable resposta en relació al grup al que pertanyen.

### 4.2.1. Variable Killed i transformació

La variable **KILLED**, tal i com s'observa a la Figura 2, té una distribució molt asimètrica, un rang de valors comprés entre 0 i 2.000, i el valor més freqüent que presenta és el zero. La mitjana és 6,57 i la desviació tipus és 60,94. Per tant, hi ha molta dispersió, ja que la desviació és gran. S'ha fet el logaritme neperià de les dades, després de sumar-les-hi 0,1 per evitar la indeterminació del  $\ln(0)$ . Aquesta transformació s'ha fet perquè com **KILLED** és una variable discreta que compta el nombre de morts en accidents industrials, i la seva variabilitat depèn del seu valor esperat, com per una Poisson, la transformació linealitzza els models que es faran posteriorment.

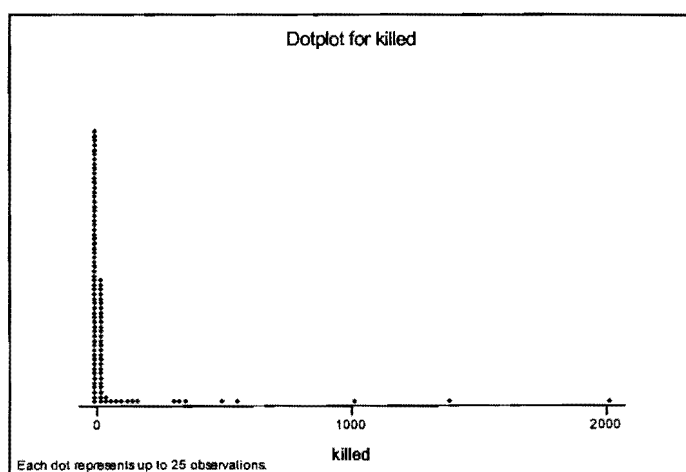


Figura 2: Diagrama de Punts de la variable **KILLED**.

A la Figura 3, hi ha representat el logaritme neperià de la variable **KILLED**. S'observa com ha disminuït la dispersió, té un rang de valors comprés entre  $-2,3026$  i  $7,6010$ , el valor més freqüent que presenta és el  $-2,3026$  que és el  $\ln(0,1)$ . La mitjana és  $-0,4767$  i la desviació tipus és  $1,9579$ .

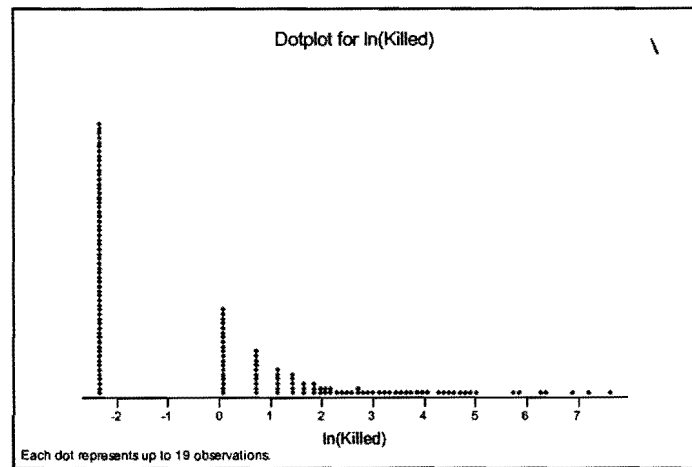
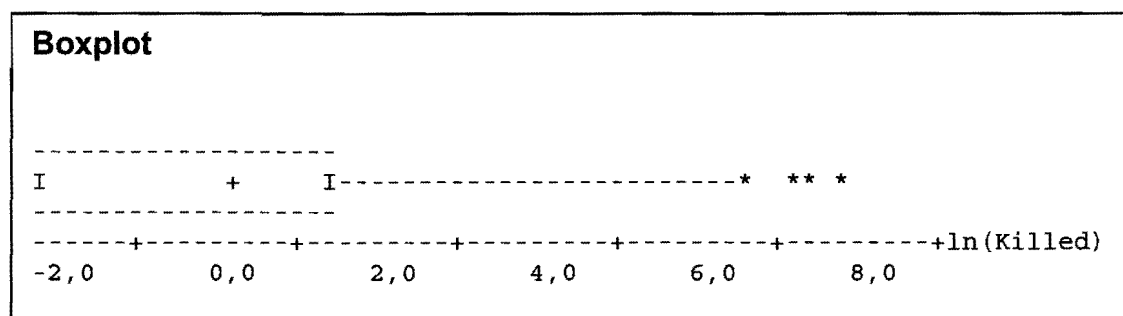


Figura 3: Diagrama de Punts del LN(KILLED).

A la Figura 4 s'hi pot observar un boxplot amb una distribució asimètrica, degut al gran nombre de zeros que té la variable **KILLED**, és a dir, que hi ha molts accidents on no hi ha cap mort. La mediana està al voltant d'un mort, el que ens indica que en la meitat dels accidents hi ha hagut 0 ó 1 (exponencial de 0,0953) mort. També s'hi pot observar la presència de 4 outliers suaus, que corresponen a 561, 1.000, 1.377 i 2.000 morts i als AN 1668, 1886, 2091 i 1098, respectivament. Aquests accidents tenen un nombre de morts més gran en comparació amb la resta, però són dades reals (no són cap error de codificació).

Figura 4: Boxplot del logaritme neperià de la variable **KILLED**.



### 4.2.2. Variable Injured i transformació

La variable **INJURED**, tal i com s'observa a la Figura 5, té una distribució molt asimètrica, un rang de valors comprès entre 0 i 9.999, i el valor més freqüent que presenta és el zero. La mitjana és 30,38 i la desviació tipus és 310,46. Per tant, hi ha molta dispersió. S'ha fet el logaritme neperià de les dades després de sumar-les-hi 0,1, pels mateixos motius que per la variable **KILLED**.

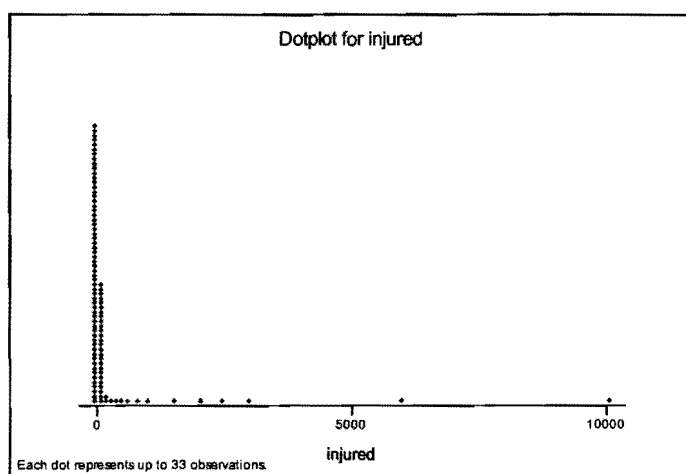
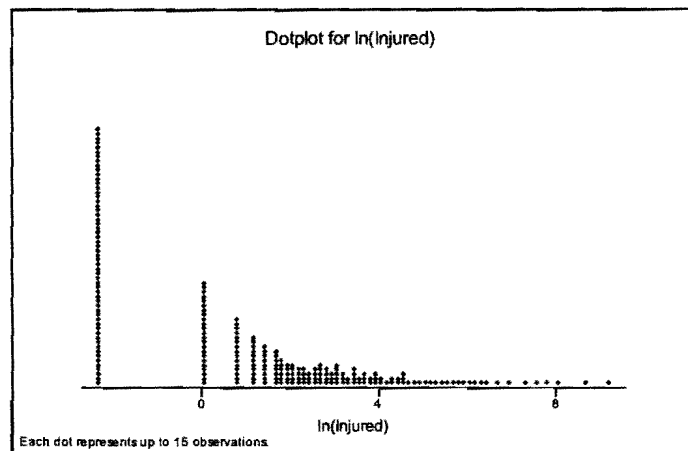
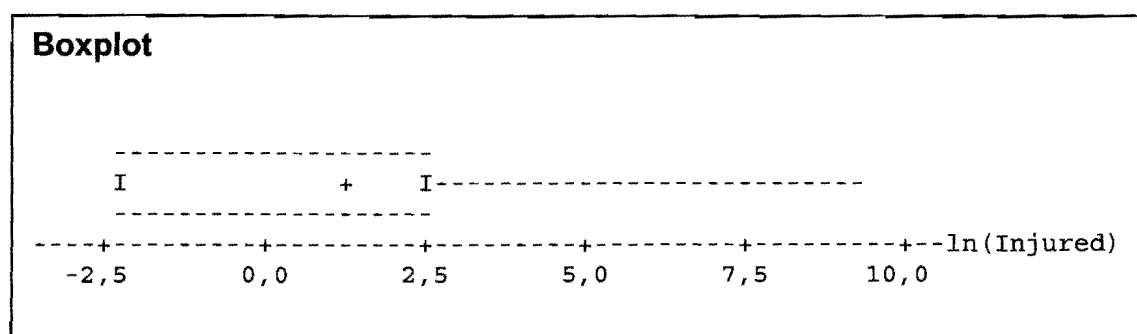


Figura 5: Diagrama de Punts de la variable **INJURED**.

A la Figura 6, hi ha representat el logaritme neperià de la variable **INJURED**. S'observa com ha disminuït la seva dispersió, ja que té un rang de valors comprès entre -2,3026 i 9,2103, el valor més freqüent que presenta és el -2,3026 que és el  $\ln(0,1)$ . La mitjana és 0,8017 i la desviació tipus és 2,3002.

Figura 6: Diagrama de Punts del  $\text{LN}(\text{INJURED})$ .

A la Figura 7 s'hi pot observar el boxplot del logaritme neperià de la variable **INJURED**, on es torna a observar que la seva distribució és asimètrica degut al gran nombre de zeros que hi ha a la variable **INJURED**, tal com també passa amb la variable **KILLED**. La mediana està al voltant de 1,13, que correspon a tres ferits, el que ens indica que en la meitat dels accidents hi ha hagut entre 0 i 3 ferits.

Figura 7: Boxplot del logaritme neperià de la variable **INJURED**.

### 4.2.3. Variable Evacuated i transformació

La variable **EVACUATED**, tal i com s'observa a la Figura 8, té un rang de valors comprés entre 0 i 200.000, i el valor més freqüent que presenta és el zero. La mitjana és 1.531 i la desviació tipus és 11.365. Com que és una variable discreta, es fa el mateix que s'ha fet amb les dues variables resposta anteriors (**KILLED** i **INJURED**), es transforma fent servir el logaritme neperià de les dades més 0,1.

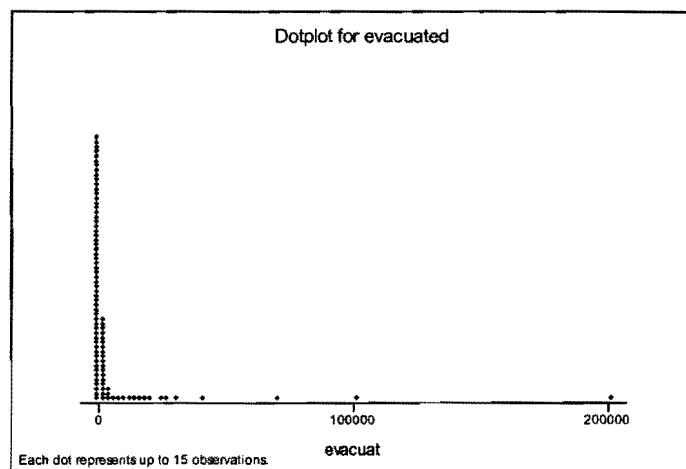


Figura 8: Diagrama de Punts de la variable **EVACUATED**.

A la Figura 9, hi ha representat el logaritme neperià de la variable **EVACUATED**. S'observa com ha disminuït la dispersió, té un rang de valors comprés entre  $-2,3026$  i  $12,206$ , i el valor més freqüent que presenta és el  $-2,3026$  que és el  $\ln(0,1)$ . La mitjana és 2,495 i la desviació tipus és 3,817. Tot i que es fa la transformació logarítmica no se solventa el problema de la no normalitat, ja que es tracta d'una variable discreta i massa asimètrica. A més a més, presenta molts missings i té molt soroll. Per aquests motius es descarta la variable **EVACUATED** com a resposta.

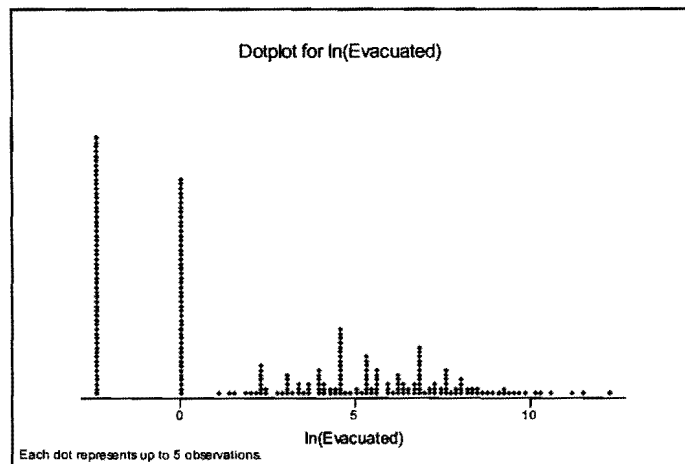


Figura 9: Diagrama de Punts del LN(EVACUATED).

A la Figura 10 s'hi pot observar un boxplot gairebé simètric, encara que una mica aplanat a la dreta igual que el dotplot anterior. La mediana està al voltant de 2,45 que correspon a 11 evacuats, el que indica que en la meitat dels accidents hi ha hagut entre 0 i 11 evacuats.

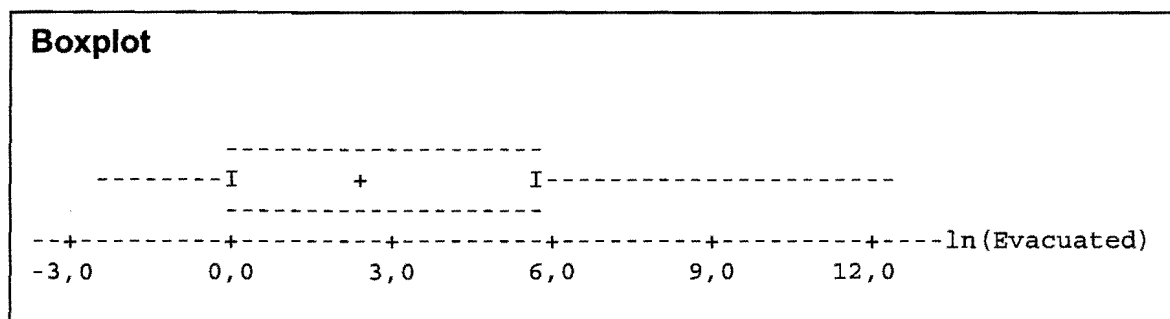


Figura 10: Boxplot del logaritme neperià de la variable EVACUATED.

#### 4.2.4. Variable danys econòmics i transformació

La variable **DAMAGE**, estimació econòmica del dany material actualitzat amb l'índex Marshall & Stevens, tal i com s'observa a la Figura 11, té un rang de valors comprés entre 0,002 i 456,53 milions de dòlars US, i el valor més freqüent és al voltant de 3,7. La mitjana és 17,53 i la desviació tipus és 46,91. Aquí es torna a fer el mateix que s'ha fet amb les tres variables anteriors (**KILLED**, **INJURED** i **EVACUATED**), el logaritme neperià de les dades més 0,1 per linealitzar el model i normalitzar la variable.

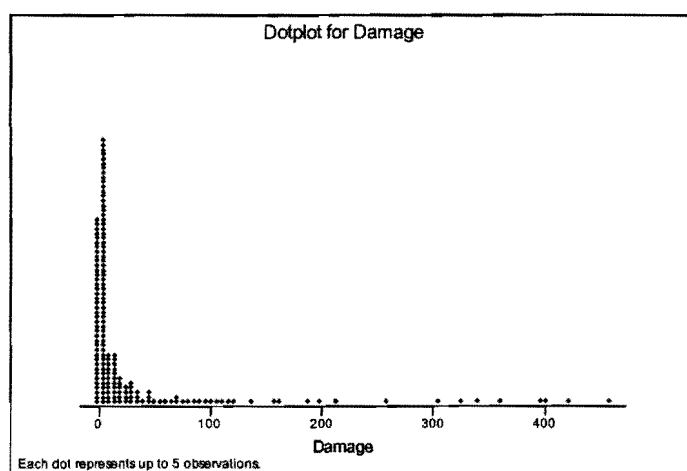
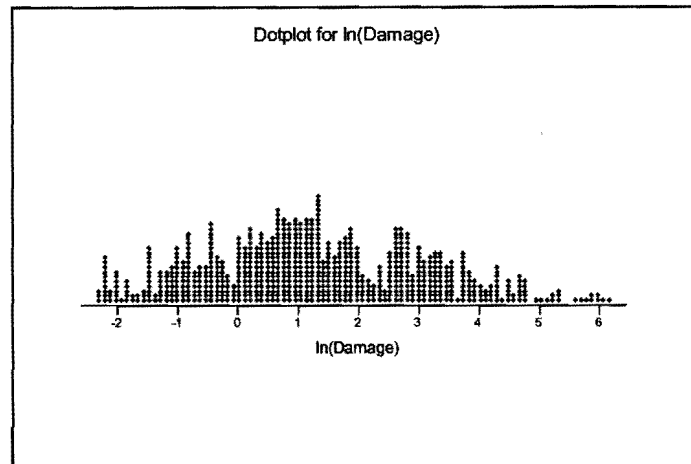
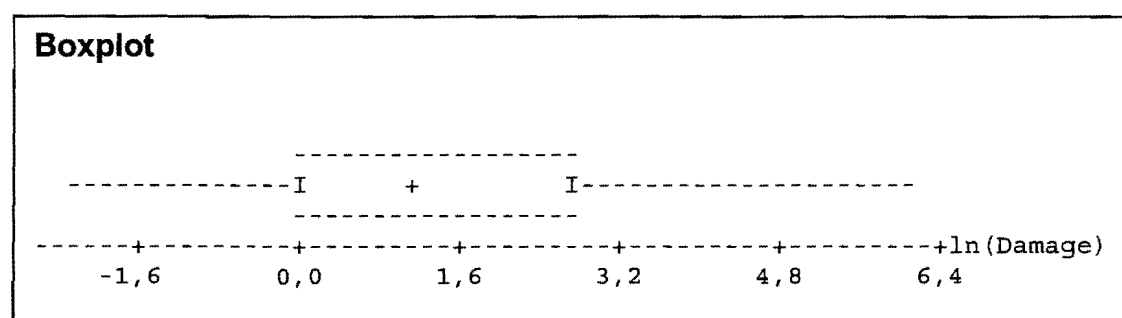


Figura 11: Diagrama de Punts de la variable **DAMAGE**.

A la Figura 12, hi ha representat el logaritme neperià de la variable **DAMAGE**. S'observa com ha disminuït la dispersió, té un rang de valors comprés entre -2,2865 i 6,1239, el valor més freqüent està al voltant de 1,35, que correspon a 3,7 milions de dòlars. La mitjana és 1,2878 i la desviació tipus és 1,7962. De les variables que s'han vist fins ara (**KILLED**, **INJURED**, **EVACUATED** i **DAMAGE**), aquesta última té la distribució més normal, ja que es tracta d'una variable continua i les altres són discretes.

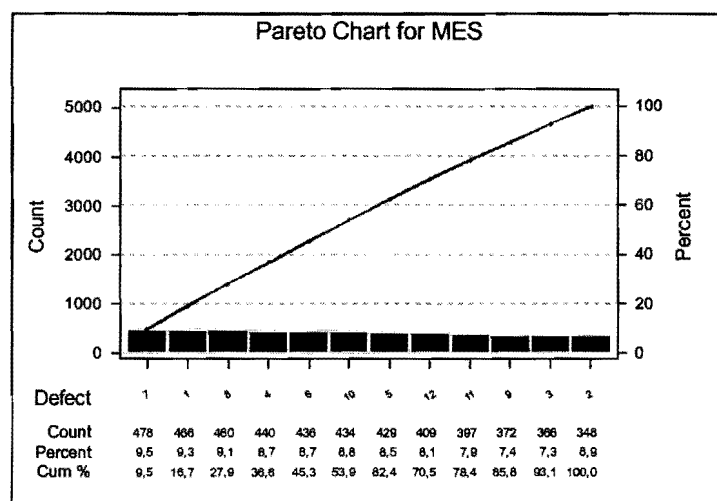
Figura 12: Diagrama de Punts del **LN(DAMAGE)**.

A la Figura 13 s'hi pot observar un boxplot gairebé simètric a l'igual que el dotplot anterior. La mediana està al voltant de 1,16, que correspon a 3,1 milions de dòlars, el que indica que la meitat dels accidents han tingut un cost del dany material entre 0,002 i 3,1 milions de dòlars US. Aquí prendre logaritmes permet aconseguir una distribució molt simètrica per primera vegada.

Figura 13: Boxplot del logaritme neperià de la variable **DAMAGE**.

## 4.3. Anàlisi univariant de les variables explicatives

### 4.3.1. Variables relacionades amb la data de l'accident



A la Figura 14 s'hi pot observar com en la variable MES, gairebé tots els mesos tenen la mateixa freqüència, no hi ha cap mes que destaquï per tenir un gran nombre d'accidents. És a dir, que presenta una distribució uniforme.

Figura 14: Diagrama de Pareto de la variable MES.

A les Figures 15, 16 i 17 s'hi pot observar que la majoria dels accidents s'han produït a finals dels anys 80 i en la dècada dels 90. Amb aquestes dues dècades es recull el 74,5% de tots els accidents. Si a aquests s'hi afegixen els accidents produïts durant els anys 70 el total d'accidents puja a un 90,3% dels accidents. Cal remarcar que aquest fet és degut a que com més recent són els accidents més informació es té sobre ells, és a dir, dels accidents produïts en la primera meitat del segle XX només es coneixen, perquè només s'han registrat, els accidents més importants. També influeix que les dades es van començar a recollir a principis dels anys 80 i que el nombre d'indústries ha augmentat molt a l'anar avançant aquest segle. Cal destacar a més a més, que en l'any 1.979 s'hi van produir un 4% del total dels accidents, un percentatge considerable.

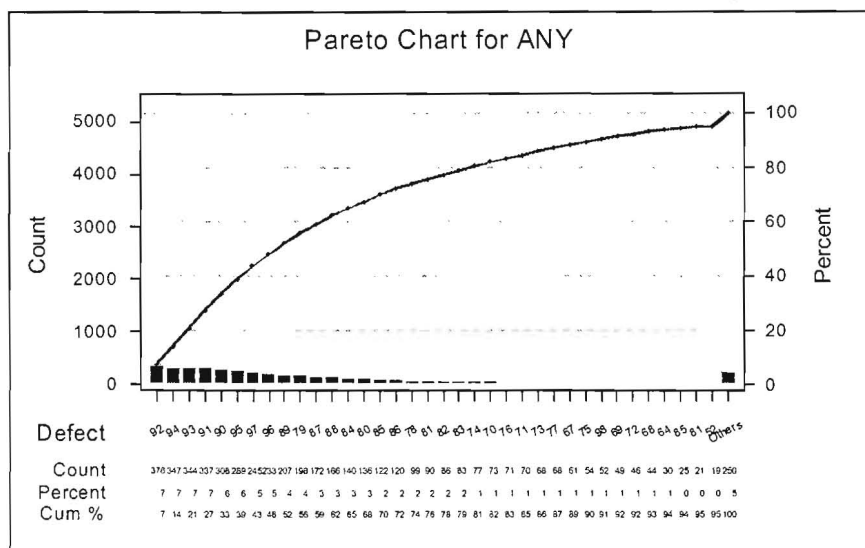


Figura 15: Diagrama de Pareto de la variable ANY.

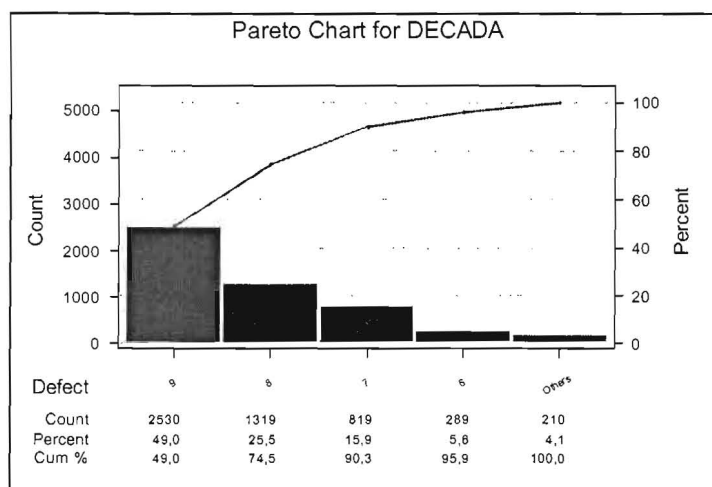


Figura 16: Diagrama de Pareto de la variable DÈCADA.

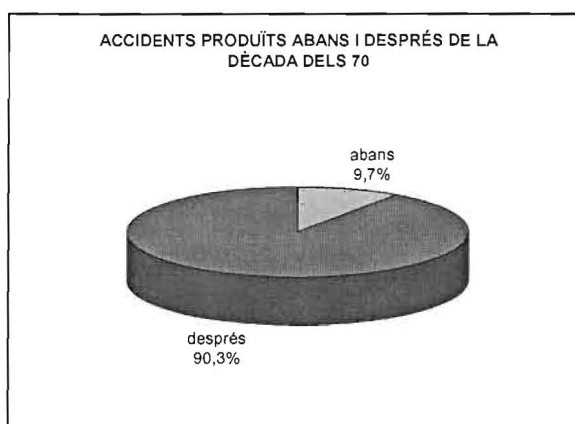
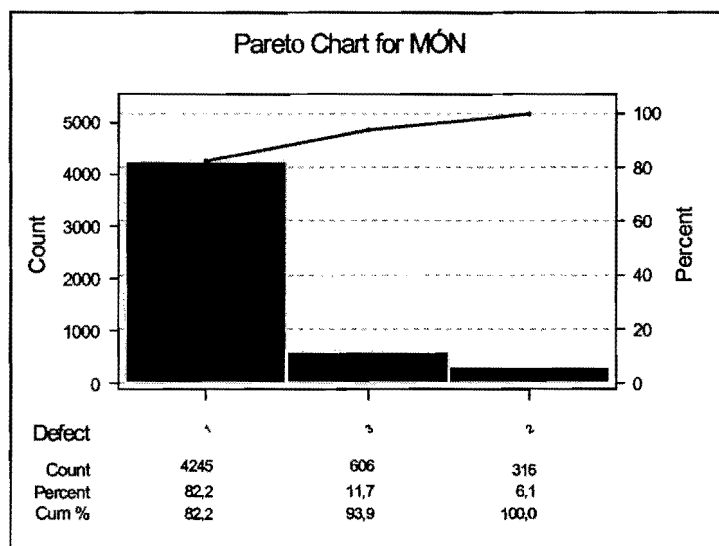


Figura 17: Diagrama de pastís de la variable ANTIGUITAT.



### 4.3.2. Variables geogràfiques

A la Figura 18 s'hi pot observar la variable que classifica els accidents segons el grau de desenvolupament del país on es van produir, **MÓN**. S'observa que un 82,2% dels accidents s'han produït en països desenvolupats. Segueixen els accidents produïts en països subdesenvolupats, amb un 11,7%, i per últim es troben els



accidents produïts en països en vies de desenvolupament que corresponen a un 6,1% dels mateixos. Ajuntant països desenvolupats i països subdesenvolupats s'agrupen gairebé un 95% de tots els accidents.

Figura 18: Diagrama de Pareto de la variable **MÓN**.

### 4.3.3. Variables que indiquen les substàncies involucrades

A la Figura 19 s'hi pot observar que l'estat físic de les substàncies que més predomina en els accidents que es coneixen és el líquid, amb un percentatge del 46%, a continuació li segueix el gas, amb un 16%, amb aquests dos estats ja es recull més de la meitat dels estats físics de les substàncies que han intervingut en un accident. També cal destacar l'estat físic gas pressuritzat amb un 11%, i el sòlid, amb un 9%. Amb tots aquests s'agrupen un 81% del total dels estats físics de les substàncies que han participat en un accident.

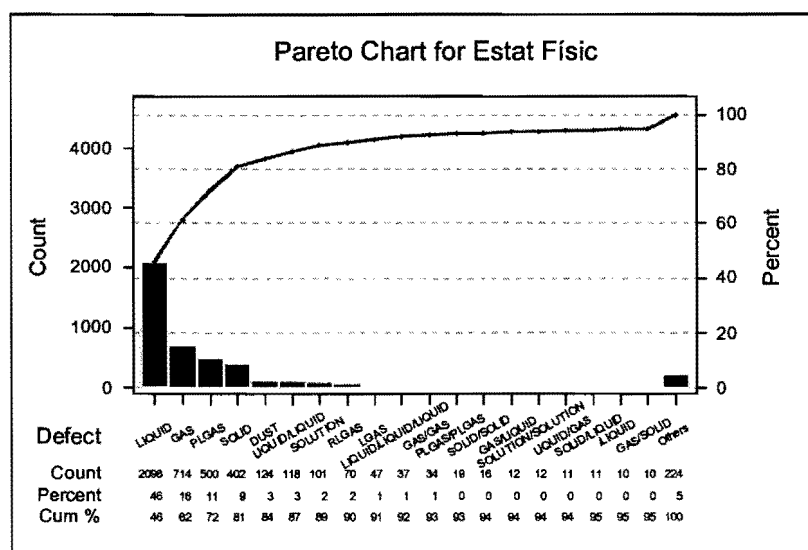


Figura 19 Diagrama de Pareto de la variable MT.

A la Figura 20 s'hi pot observar la variable estat físic de les substàncies amb la nova categorització (veure apartat 3.3.3) que consisteix en agrupar els estats físics de les substàncies en 5 categories: SÒLID, LÍQUID, GAS, PLGAS i DUST. Aquesta és la categorització que es farà servir en el modelat de les variables resposta.

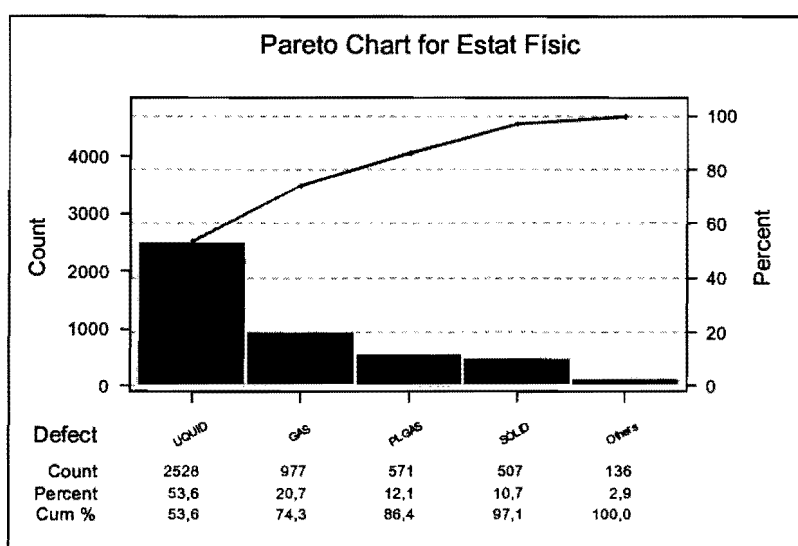
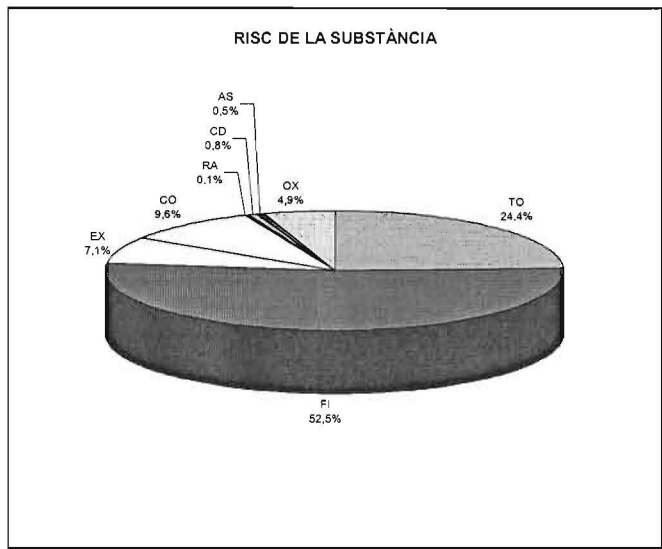


Figura 20: Diagrama de Pareto de la variable ESTAT\_FÍSIC.

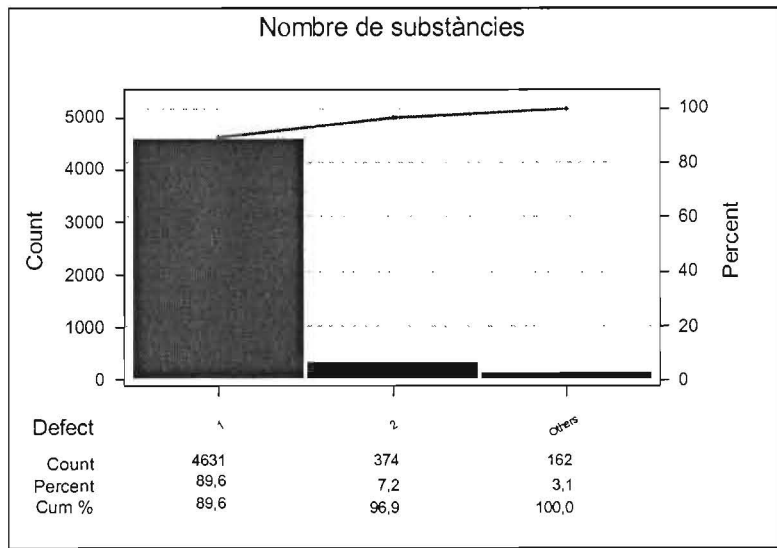
A la Figura 21 s’hi pot observar les categories que indiquen el risc de les substàncies que intervenen en els accidents. Més de la meitat de les substàncies que intervenen en accidents són inflamables, FI. A continuació es troben les substàncies tòxiques, TO, amb un 24,4%, les substàncies corrosives, CO, amb un 9,6% i les



explosives, EX, amb un 7,1%. Amb tots ells es recull un 93,6% de tots els riscos de les substàncies que han intervingut en accidents. Cal recordar que aquestes categories no són excloents i els percentatges no són sobre el total d’accidents.

Figura 21: Diagrama de Pastís de la variable **HAZARD**.

A la Figura 22 s’hi pot observar, com ja s’havia mencionat que, en la majoria d’accidents només hi intervé una substància, 89,6%. El percentatge d’accidents en



els que intervé dues substàncies és del 7,2%, i a l’augmentar el nombre de substàncies el percentatge encara es redueix més.

Figura 22 Diagrama de Pareto de la variable **NSUS**.

#### 4.3.4. Variables que descriuen el tipus d'accident

A la Figura 23 s'hi pot observar que un 29,9% dels accidents s'han produït per una fuga, RELEASE. A continuació li segueixen els accidents que s'han produït per un incendi, FIRE, amb un 19,9%. També tenen un percentatge important els accidents que s'han produït per una explosió que ha provocat posteriorment un incendi, EXPLODE-FIRE, 13,7%. Els accidents que s'han produït per una explosió, EXPLODE, representen el 11,8% sobre el total. Per finalitzar es destaquen els accidents que s'han produït per una fuga que posteriorment ha originat un núvol de gas, RELEASE-GASCLD, amb un 6,5%.

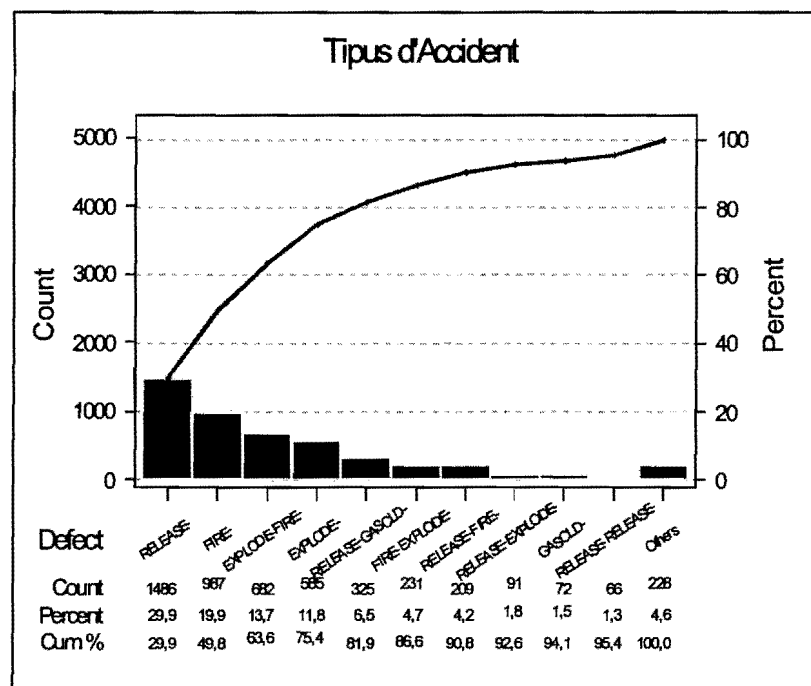
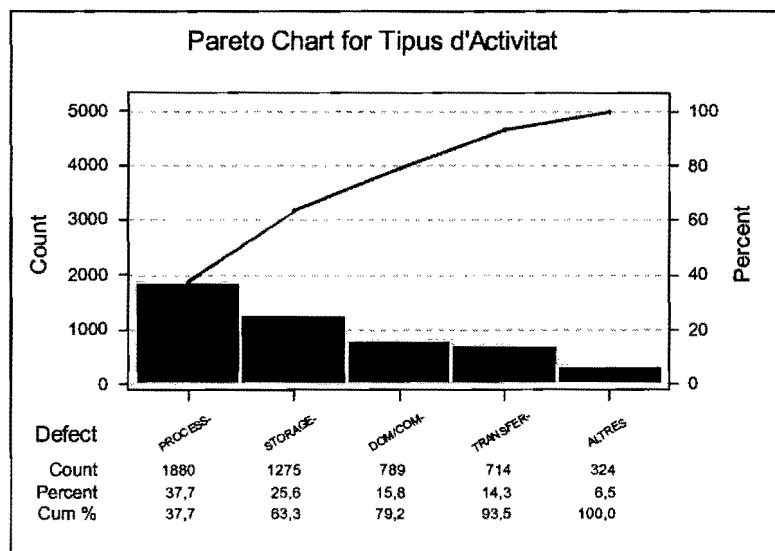


Figura 23: Diagrama de Pareto de la variable **TIPUSACC**.

### 4.3.5. Variables que expliquen les causes de l'accident

A la Figura 24 s'hi pot observar que el tipus d'activitat més freqüent és quan hi ha hagut un accident en una planta de procés, PROCESS. Aquests són un 37,7% dels accidents. A continuació es troben els accidents originats en unitats o àrees d'una planta d'emmagatzematge, STORAGE, amb un 25,6%; els accidents produïts



en locals domèstics o comercials, DOM/COM, amb un 15,8% i els accidents produïts durant operacions de càrrega i descàrrega, TRANSFER, amb un 14,3%.

Figura 24: Diagrama de Pareto de la variable TIPUSACT.

A la Figura 25 s'hi pot observar que la causa general de l'accident més freqüent és per una fallida mecànica, MECHANICAL, amb un 34,8%. A continuació es troba la causa humana, HUMAN, amb un 27,1%, li segueix amb un percentatge del 19,4% els accidents produïts per un esdeveniment extern, EXTERNAL, com per exemple un huracà. Després es troben els accidents produïts per una reacció violenta, VREACTION, amb un 8,6%. També els accidents produïts per impacte, IMPACT, tenen un percentatge semblant, un 5,2%. Amb totes aquestes causes es recull les causes generals del 95,2% dels accidents. Cal remarcar que aquestes categories no són excloents i pot ser que un accident fos produït, per exemple, per una errada humana i mecànica a l'hora.

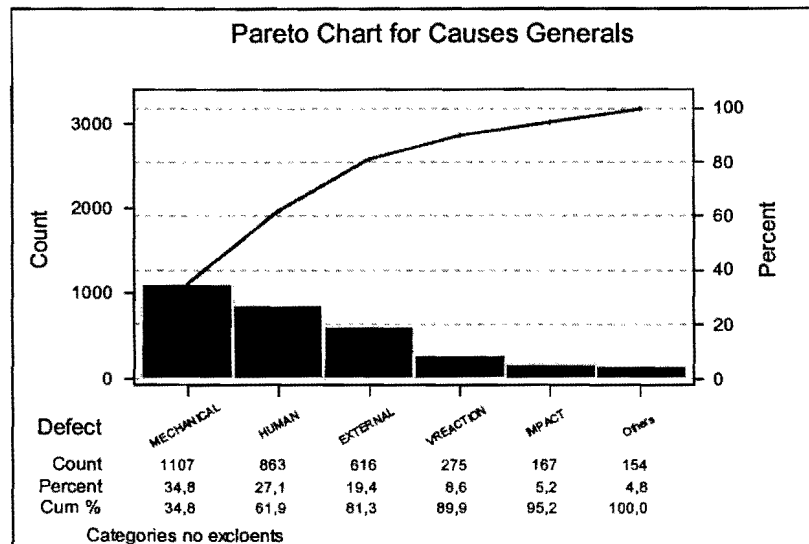
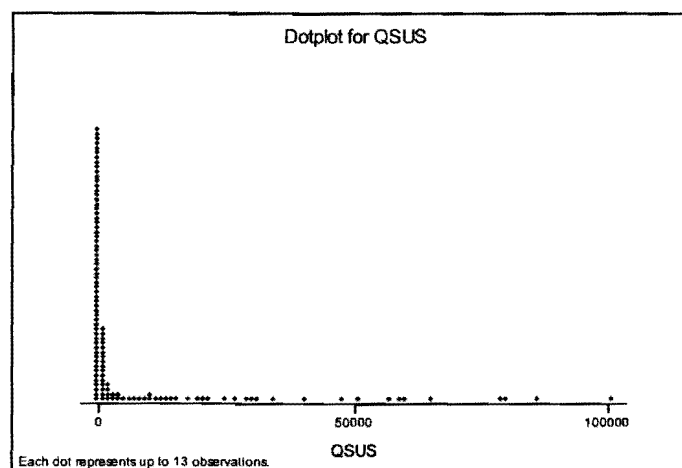


Figura 25: Diagrama de Pareto de la variable CAUSAG.

#### 4.3.6. Variable que mesura la quantitat de substància involucrada i transformació

A la Figura 26 s'hi pot observar com la variable QSUS, quantitat de substància que ha intervingut en l'accident en tones, té un rang de valors comprés entre 0 i 99.999, i el valor més freqüent que presenta és 1. La mitjana és 2.156 i la desviació



tipus és 9.710. Per tant, hi ha molta dispersió, ja que la desviació és molt gran. Per reduir la variabilitat i normalitzar la variable s'ha pres el logaritme neperià de les dades després de sumar-les-hi 0,1.

Figura 26: Diagrama de Punts de la variable QSUS.

A la Figura 27 està representat el logaritme neperià de la variable **QSUS**. S'observa com ha disminuït la dispersió, té un rang de valors comprès entre  $-2,3026$  i  $11,5129$ , el valor més freqüent que presenta és el  $0,0953$  que és el  $\ln(1,1)$ . La mitjana és  $3,1520$  i la desviació tipus és  $3,3363$ . També s'observa que té una distribució bastant normal.

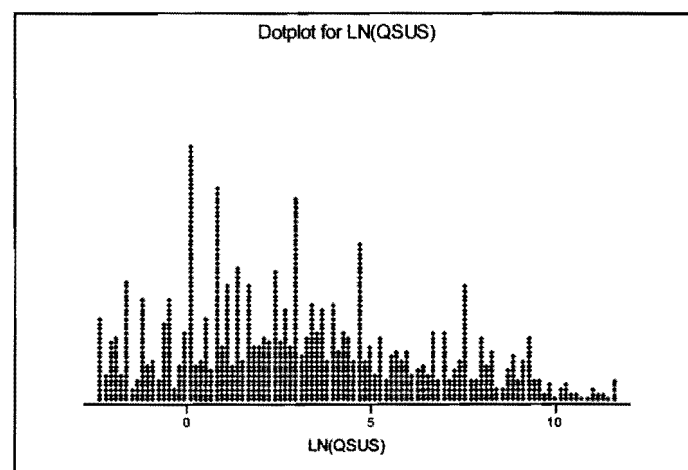


Figura 27: Diagrama de Punts del LN(QSUS).

A la Figura 28 s'hi pot observar un boxplot simètric igual que el dotplot anterior, a diferència del dotplot inicial que és asimètric (Figura 26). La mediana està al voltant de  $2,84$ , que correspon a 17 tones, el que indica que en la meitat dels accidents han participat de 0 a 17 tones.

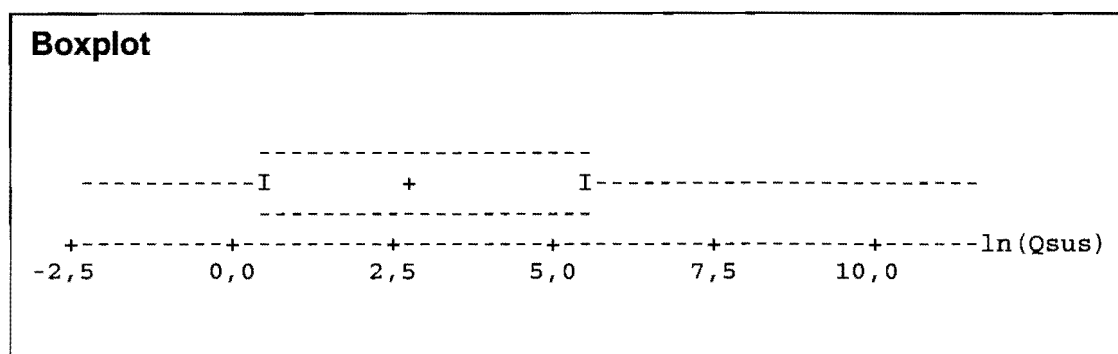
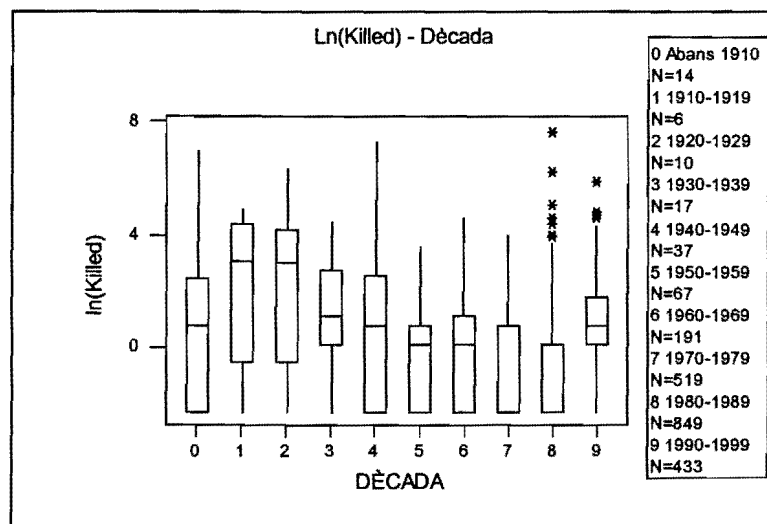


Figura 28: Boxplot del logaritme neperià de la variable **QSUS**.

## 4.4. Anàlisi bivariant per Killed

### 4.4.1. Relació de Killed amb la data de l'accident

A la Figura 29 s'hi pot observar el logaritme neperià de la variable **KILLED** segons la dècada en què es va produir l'accident. No hi ha hagut cap dècada que tingui un boxplot simètric. El nombre de morts varia segons la dècada en què es produeix l'accident. En general en totes les dècades hi ha molta dispersió, la que en té menys és la dècada dels 90. Exceptuant els accidents anteriors al 1910, el nombre



de morts disminueix fins arribar a la dècada dels noranta que augmenta respecte a la dècada dels vuitanta. El nombre de morts per les dos primeres dècades és semblant i elevat.

Figura 29: Boxplot múltiple LN(KILLED) vs DÈCADA.

Per la dècada dels 80 destaquen els accidents següents per tenir un nombre elevat de morts: AN 420 amb 500 morts, AN 989 amb 153 morts, AN 1098 amb 2.000 morts, AN 1376 amb 80 morts, AN 2228 amb 54 morts, AN 2317 amb 51 morts i AN 2954 amb 100 morts.



I per la dècada dels 90: **AN** 5150 amb 110 morts, **AN** 5219 amb 120 morts, **AN** 7861 amb 100 morts i **AN** 8418 amb 343 morts.

Consultant amb l'expert químic responsable del projecte, s'ha comprovat que aquests outliers són dades reals i no corresponen a cap error a l'introduir les dades.

Tal i com s'observa a la Taula 4, a la dècada dels anys 20 hi ha hagut un major nombre de morts en mitjana i en la dècada dels anys 50 on hi ha hagut menys víctimes mortals i menys desviació. Els accidents anteriors al 1.909 (aquest inclòs, codificats amb 0) són en canvi els que tenen una desviació tipus més gran. Cal destacar també que a partir de la dècada dels 50 el nombre de morts en mitjana disminueix bastant.

| Dècada | N   | N <sup>*1</sup> | Mean  | Median | StDev  | Minimum | Maximum |
|--------|-----|-----------------|-------|--------|--------|---------|---------|
| 0      | 14  | 1               | 81,8  | 2      | 265,6  | 0       | 1.000   |
| 1      | 6   | 0               | 42,3  | 25,5   | 51,7   | 0       | 133     |
| 2      | 10  | 1               | 97,5  | 20     | 186,4  | 0       | 561     |
| 3      | 17  | 3               | 14,65 | 3      | 25,16  | 0       | 82      |
| 4      | 37  | 13              | 55,9  | 2      | 230,3  | 0       | 1.377   |
| 5      | 67  | 41              | 2,522 | 1      | 5,714  | 0       | 35      |
| 6      | 191 | 98              | 3,246 | 1      | 10,061 | 0       | 100     |
| 7      | 519 | 300             | 2,669 | 0      | 6,933  | 0       | 52      |
| 8      | 849 | 470             | 4,88  | 0      | 71,12  | 0       | 2.000   |
| 9      | 433 | 2.097           | 7,125 | 2      | 20,463 | 0       | 343     |

Taula 4: Descriptiva de la variable **KILLED** vs la **DÈCADA**.

A la Figura 30 s'hi pot observar el logaritme neperià de la variable **KILLED** segons si l'accident ha succeït abans o després de la dècada dels 70. Tant en el boxplot d'abans dels 70 com després dels 70 hi ha un outlier, per tenir un nombre elevat de morts.

Abans dels 70 destaca l'accident amb **AN** 2091 amb 1.377 morts i després dels 70 destaca l' accident amb **AN** 1098 amb 2.000 morts (aquest accident ja s'havia registrat abans, a la Figura 26, com dada atípica). Els dos boxplots són asimètrics.

---

<sup>1</sup> Nombre de missings.

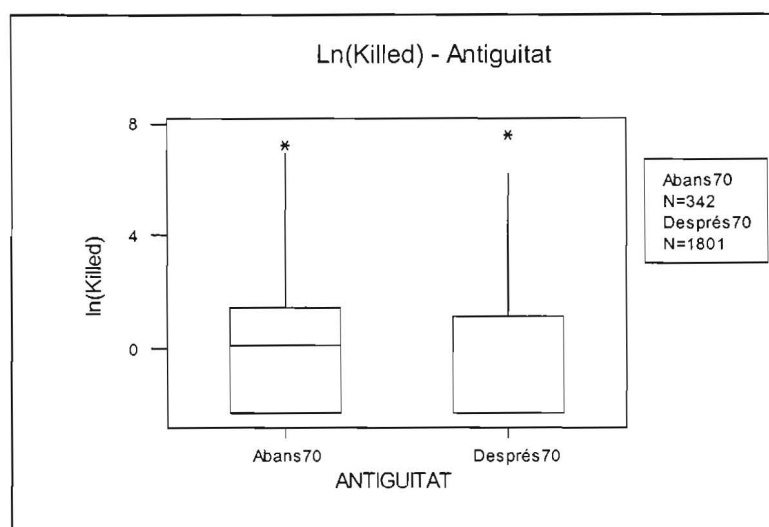


Figura 30: Boxplot múltiple del **LN(KILLED)** vs **ANTIGUITAT**.

Els accidents que es van produir abans de la dècada dels setanta tenen 16,02 morts de mitjana amb una desviació tipus de 100,11. I els accidents que es van produir després de la dècada dels setanta tenen 4,78 morts de mitjana amb una desviació tipus de 50. Cal destacar que a partir de la dècada dels anys 70 el nombre de morts ha disminuït bastant, el mateix passa amb la variabilitat, però encara és gran.

| Antiguitat      | N     | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|-----------------|-------|-------|-------|--------|--------|---------|---------|
| 0 (Abans 70)    | 342   | 157   | 16,02 | 1      | 100,11 | 0       | 1.377   |
| 1 (A partir 70) | 1.801 | 2.867 | 4,78  | 0      | 50     | 0       | 2.000   |

Taula 5: Descriptiva de la variable **KILLED** vs **ANTIGUITAT**.

#### 4.4.2. Relació de Killed amb la situació geogràfica

A la Figura 31 està representat el logaritme neperià de la variable **KILLED** segons el grau de desenvolupament del país on va succeir l'accident. El nombre de morts dels accidents produïts en estats en via de desenvolupament i subdesenvolupats s'assembla molt i és diferent en comparació amb els estats desenvolupats on en general hi ha menys.

Destaquen diverses observacions atípiques per tenir un nombre elevat de morts.

Els accidents atípics que es van produir en estats desenvolupats són: AN 1668 amb 561 morts, AN 1886 amb 1.000 morts i AN 2416 amb 321 morts.

Només hi ha un accident atípic que es va produir en estats en via de desenvolupament, aquest té com AN el 420 i es van produir 500 morts.

Els accidents atípics que es van produir en estats subdesenvolupats són: AN 1098 amb 2.000 morts, AN 2091 amb 1.377 morts i AN 8418 amb 343 morts.

Cal destacar que els accidents atípics produïts en països en via de desenvolupament o subdesenvolupats ja havien sobresortit de la resta per tenir un gran nombre de morts, en gràfics anteriors.

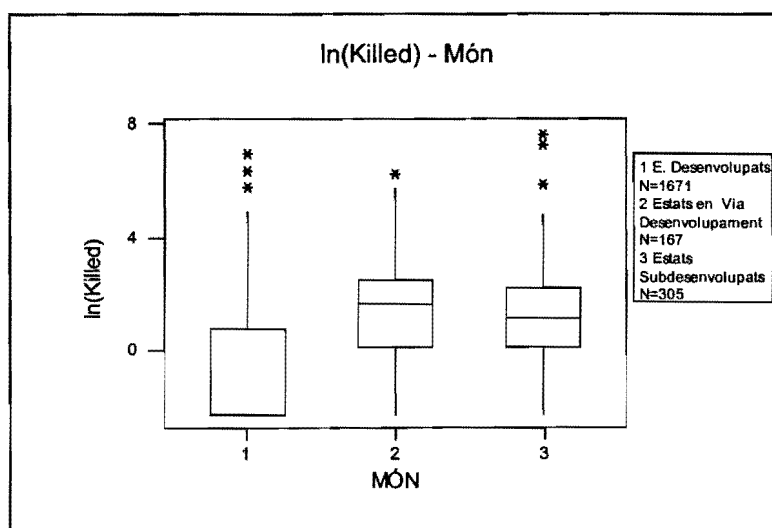


Figura 31: Boxplot múltiple del LN(KILLED) vs MÓN.

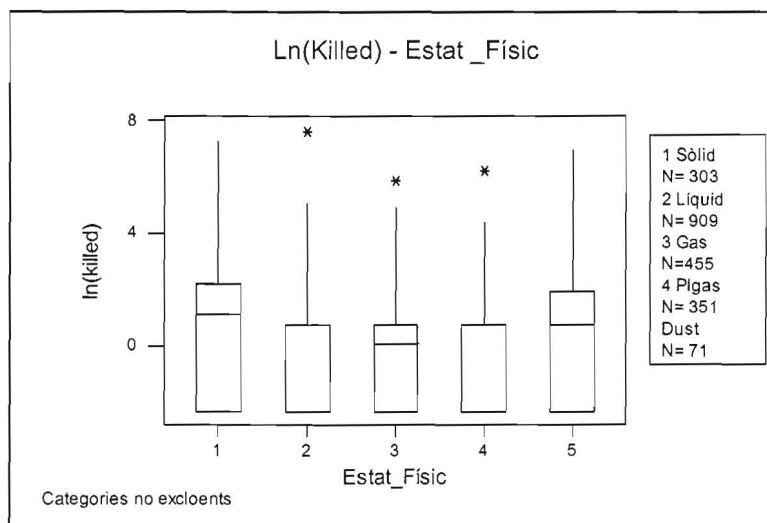
Tal i com s'observa a la Taula 6, en mitjana els accidents que causen més morts són els que es produeixen en països subdesenvolupats i són els que tenen una dispersió més gran. Per contra, els accidents que tenen menys morts i menys variabilitat són els que es produeixen en estats desenvolupats.

| Món                              | N     | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|----------------------------------|-------|-------|-------|--------|--------|---------|---------|
| 1 (Estats Desenvolupats)         | 1.671 | 2.574 | 3,18  | 0      | 29,98  | 0       | 1.000   |
| 2 (E. en via de Desenvolupament) | 167   | 149   | 14,92 | 5      | 47,4   | 0       | 500     |
| 3 (E. Subdesenvolupats)          | 305   | 301   | 20,61 | 3      | 140,36 | 0       | 2.000   |

Taula 6: Descriptiva de la variable **KILLED** vs **MÓN**.

### 4.4.3. Relació de Killed amb les substàncies involucrades

A la Figura 32 s'hi pot observar el logaritme neperià de la variable **KILLED** en funció de l'estat físic de la substància o substàncies que han intervingut en l'accident. El nombre de morts que causen els accidents on participen substàncies en estat físic sòlid o pols, DUST, són similars, i el nombre de morts que provoquen



els accidents on intervenen substàncies en estat físic líquid i gas pressuritzat, PLGAS, són semblants.

Figura 32: Boxplot múltiple **LN(KILLED)** vs **ESTAT\_FÍSIC**.

En els accidents on han participat substàncies en estat físic líquid destaca una observació atípica per tenir un nombre de morts més alt que la resta d'accidents. Aquest accident té **AN** 1098 i 2.000 morts. En els accidents on han participat substàncies en estat físic gas destaca una observació atípica per tenir un nombre de morts més alt que la resta d'accidents. Aquest accident té **AN** 8418 i 343 morts. I per últim també cal destacar d'entre els accidents on han intervingut substàncies en estat físic gas pressuritzat l'accident amb **AN** 420 per tenir 500 morts.

Tal i com s'observa a la Taula 7, els accidents que tenen més morts en mitjana i més variabilitat són aquells en els que intervenen substàncies en estat físic pols i en estat físic sòlid. En canvi, els accidents que tenen menys morts en mitjana i menys variabilitat són aquells en els que intervenen substàncies en estat físic gas, seguits de les substàncies en estat físic gas pressuritzat (si no es té en compte els missings).

| Estat_físic          | N   | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|----------------------|-----|-------|-------|--------|--------|---------|---------|
| 1 (Sòlid)            | 303 | 204   | 18,15 | 3      | 90,13  | 0       | 1.377   |
| 2 (Líquid)           | 909 | 1.619 | 4,17  | 0      | 66,64  | 0       | 2.000   |
| 3 (Gas)              | 455 | 522   | 3,765 | 1      | 18,438 | 0       | 343     |
| 4 (Gas Pressuritzat) | 351 | 220   | 3,93  | 0      | 27,58  | 0       | 500     |
| 5 (Pols)             | 71  | 65    | 21,2  | 2      | 118,7  | 0       | 1.000   |
| *                    | 131 | 468   | 3,893 | 1      | 7,004  | 0       | 43      |

Taula 7: Descriptiva de **KILLED** vs **ESTAT\_FÍSIC**.

A la Figura 33 es mostra el logaritme neperià de la variable **KILLED** en funció del risc de la substància o substàncies que han participat en l'accident. El nombre de morts és diferent segons el risc de la substància o substàncies que hagin participat en l'accident. Les substàncies que més morts causen són les explosives i les asfixiants.

Apareixen diferents accidents atípics per tenir un nombre elevat de morts en relació als accidents on han participat substàncies:

Tòxiques: AN 1098 amb 2.000 morts, AN 1750 amb 60 morts, AN 1931 amb 82 morts i AN 5438 amb 41 morts

Inflamables: AN 420 amb 500 morts, AN 1098 amb 2.000 morts i AN 8418 amb 343 morts.

Explosives: AN 1668 amb 561 morts, AN 1886 amb 1.000 morts i AN 2091 amb 1.377 morts.

Corrosives: AN 1931 que té 82 morts.

Oxidants: AN 1668 amb 561 morts.

I finalment, destaca per no tenir cap mort en accidents on hi ha participat substàncies asfixiants l'accident amb AN 1711.

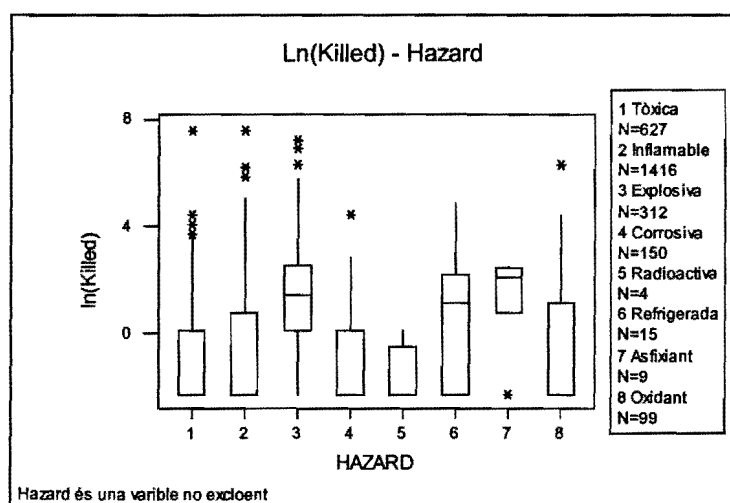


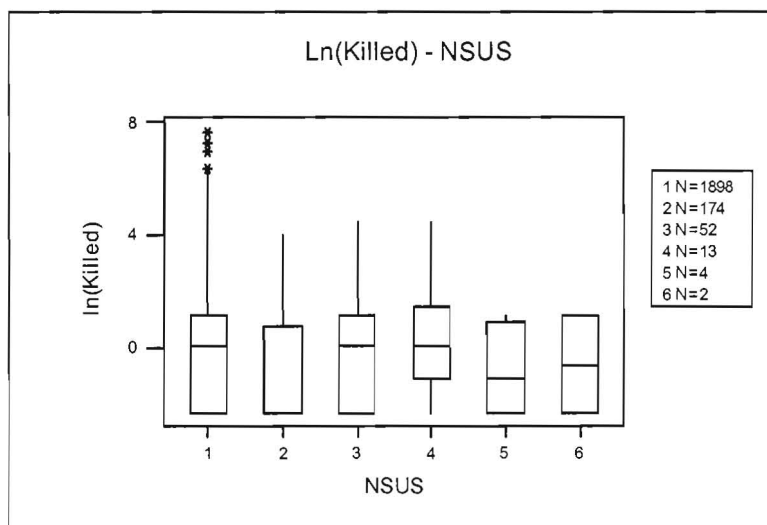
Figura 33: Boxplot múltiple del LN(KILLED) vs HAZARD.

Tal i com s'observa a la Taula 8, els accidents amb més morts en mitjana són aquells on han participat substàncies explosives, seguits dels accidents on han intervingut substàncies refrigerades. Els accidents on han participat substàncies explosives i tòxiques són els que tenen una desviació més gran en el nombre de morts. D'altra banda, els accidents que tenen menys morts en mitjana són aquells on han participat substàncies corrosives, si no es tenen en compte les substàncies radioactives. Els accidents amb menys desviació són aquells en els que han participat substàncies asfixiants, si no es tenen en compte les substàncies radioactives.

| HAZARD          | N     | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|-----------------|-------|-------|-------|--------|--------|---------|---------|
| 1 (Tòxica)      | 627   | 894   | 4,66  | 0      | 79,99  | 0       | 2.000   |
| 2 (Inflamable)  | 1.416 | 1.853 | 4,58  | 0      | 55,97  | 0       | 2.000   |
| 3 (Explosiva)   | 312   | 132   | 22,91 | 4      | 104,82 | 0       | 1.377   |
| 4 (Corrosiva)   | 150   | 449   | 1,267 | 0      | 6,903  | 0       | 82      |
| 5 (Radioactiva) | 4     | 1     | 0,25  | 0      | 0,5    | 0       | 1       |
| 6 (Refrigerada) | 15    | 32    | 15,13 | 3      | 34,4   | 0       | 131     |
| 7 (Asfixiant)   | 9     | 23    | 6,56  | 8      | 4,80   | 0       | 12      |
| 8 (Oxidant)     | 99    | 208   | 8,55  | 0      | 56,81  | 0       | 561     |

Taula 8: Descriptiva de la variable **KILLED** vs **HAZARD**.

A la Figura 34 està representat el logaritme neperià de la variable **KILLED** en



funció del nombre de substàncies que han intervingut en l'accident. Els accidents amb més morts són aquells on han participat 4 substàncies.

Figura 34: Boxplot múltiple del **LN(KILLED)** vs **NSUS**.

En els accident on només intervé una substància destaquen 4 observacions atípiques per tenir un nombre elevat de morts, que són: **AN 1098** amb 2.000 morts, **AN 1668** amb 561 morts, **AN 1886** amb 1.000 morts i **AN 2091** amb 1.377 morts.

Tal i com s'observa a la Taula 9, els accidents on hi ha hagut més morts en mitjana han estat aquells on han participat quatre substàncies. També cal destacar els accidents on només intervé una substància per tenir la dispersió més gran, però també és cert que són els accidents més nombrosos.

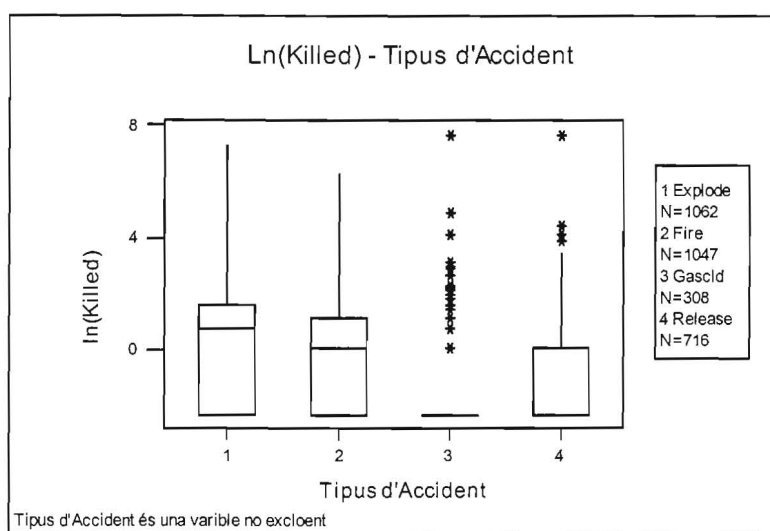
En canvi, els que han tingut menys morts en mitjana i menys variabilitat han estat els accidents on han intervingut cinc o sis substàncies, però hi ha pocs accidents. Li segueixen els accidents on intervenen dues substàncies.

| NSUS | N     | N*    | Mean  | Median | StDev | Minimum | Maximum |
|------|-------|-------|-------|--------|-------|---------|---------|
| 1    | 1.898 | 2.733 | 6,99  | 1      | 64,65 | 0       | 2.000   |
| 2    | 174   | 200   | 2,661 | 0      | 6,528 | 0       | 53      |
| 3    | 52    | 53    | 4,77  | 1      | 12,71 | 0       | 82      |
| 4    | 13    | 33    | 8,54  | 1      | 21,84 | 0       | 80      |
| 5    | 4     | 4     | 1     | 0,5    | 1,414 | 0       | 3       |
| 6    | 2     | 1     | 1,5   | 1,5    | 2,12  | 0       | 3       |

Taula 9: Descriptiva de **KILLED** vs **NSUS**.

#### 4.4.4. Relació de Killed amb el tipus d'accident

A la Figura 35 s'hi pot observar el logaritme neperià de la variable **KILLED** segons el tipus d'accident. El nombre de morts és semblant en els accidents que es produeixen per explosió, **EXPLODE**, o incendi, **FIRE**. Els accidents produïts per un núvol de gas, **GASCLD**, són menys greus i en la majoria no hi ha hagut cap mort. I



per últim els accidents produïts per una fuga, **RELEASE**, provoquen menys morts que els produïts per un incendi o una explosió.

Figura 35: Boxplot múltiple del **LN(KILLED)** vs **TIPUSACC**.



Dels accidents produïts per un núvol de gas, destaquen els accidents que han tingut com a mínim un mort. En aquest cas hi ha 235 accidents en els quals no han hagut cap mort i 73 accidents en els quals hi ha hagut com a mínim un mort. Per aquest motiu els accidents on han hagut víctimes mortals queden marcats com observacions atípiques, ja que són una minoria respecte als que no han tingut víctimes mortals.

Dels accidents produïts per una fuga, RELEASE, destaquen per tenir un nombre elevat de morts els següents accidents: **AN 534** amb 49 morts, **AN 1098** amb 2.000 morts, **AN 1750** amb 60 morts i **AN 1818** amb 80 morts.

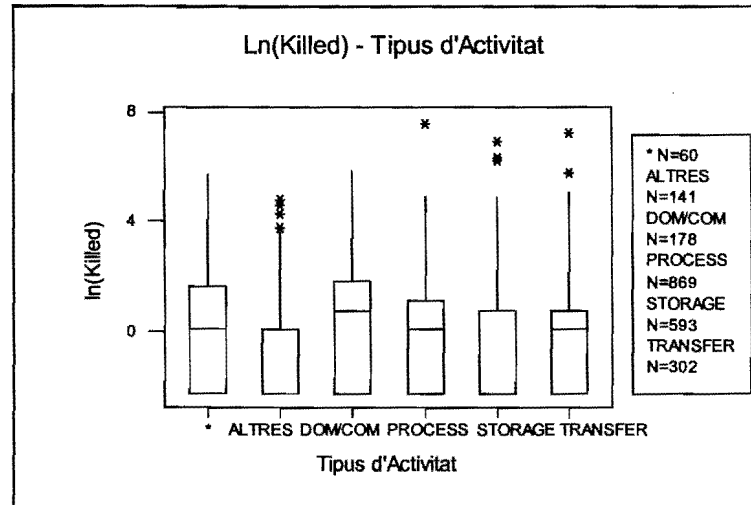
Tal i com s'observa a la Taula 10, els accidents que provoquen més morts en mitjana són els que es produeixen per explosió. En canvi, els accidents amb menys morts són els accidents produïts per una fuga seguits dels que es produeixen per un incendi amb una mitjana semblant i amb la menor variabilitat, si no es tenen en compte els missings.

| Tipus Accident   | N     | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|------------------|-------|-------|-------|--------|--------|---------|---------|
| 1 (Explosió)     | 1.062 | 613   | 9,76  | 2      | 60,69  | 0       | 1.377   |
| 2 (Incendi)      | 1.047 | 1.193 | 4,242 | 1      | 20,961 | 0       | 500     |
| 3 (Núvol de Gas) | 308   | 208   | 7,88  | 0      | 114,2  | 0       | 2.000   |
| 4 (Fuga)         | 716   | 1.550 | 4,18  | 0      | 74,87  | 0       | 2.000   |
| *                | 56    | 149   | 4     | 1      | 8,67   | 0       | 39      |

Taula 10: Descriptiva de la variable **KILLED** vs **TIPUSACC**.

#### 4.4.5. Relació de Killed amb el tipus d'activitat

La Figura 36, representa el logaritme neperià de la variable **KILLED** segons el tipus d'activitat que es duia a terme en el moment de l'accident. El nombre de morts és clarament diferent segons el tipus d'activitat. No hi ha cap tipus d'activitat que



segueixi una distribució simètrica. Els accidents produïts en activitats auxiliars de la indústria, ALTRES, és el tipus d'activitat que causa menys morts.

Figura 36: Boxplot múltiple del LN(KILLED) vs TIPUSACT.

A la categoria ALTRES destaquen quatre accidents atípics per tenir un elevat nombre de morts: AN 2954 amb 100 morts, AN 5219 amb 120 morts, AN 5522 amb 43 morts i AN 7003 amb 73 morts.

Aquest accidents no havien aparegut abans pel seu comportament atípic, a excepció dels AN 5219 i AN 2954.

Dels accidents produïts en plantes de procés, PROCESS, destaca l'accident amb AN 1098 amb 2.000 morts, aquest és el famós accident que es va produir a Bhopal i surt gairebé a tots els gràfics vists fins ara com a outlier.

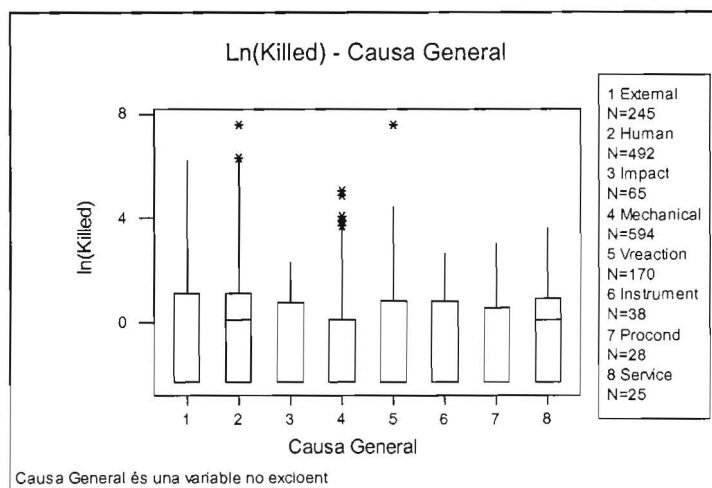
Dels accidents que es van produir en plantes d'emmagatzematge, STORAGE, destaquen: AN 420 amb 500 morts, AN 1668 amb 561 morts i AN 1886 amb 1.000 morts.

També destaquen els següents accidents produïts durant operacions de càrrega i descàrrega, DOM/COM: **AN 2091** amb 1.377 morts i **AN 2416** amb 321 morts.

Tal i com s'observa a la Taula 11, els accidents amb més morts en mitjana es produeixen en plantes de procés, seguits dels produïts en àrees d'emmagatzematge. Aquests últims són els que tenen més variabilitat. Per contra, els accidents amb menys morts i menys variabilitat són els produïts durant tasques auxiliars de la indústria, ALTRES, si no es té en compte els missings, seguits dels accidents produïts durant operacions de càrrega i descàrrega.

| TIPUS D'ACTIVITAT                             | N   | N*   | Mean  | Median | StDev | Minimum | Maximum |
|---|-----|------|-------|--------|-------|---------|---------|
| ALTRES (tasques auxiliars de la indústria)    | 79  | 245  | 9,73  | 2,71   | 18,18 | 0       | 114,97  |
| DOM/COM (locals domèstics o comercials)       | 52  | 737  | 12,59 | 1,16   | 45,28 | 0,01    | 303,17  |
| PROCESS (planta de procés)                    | 329 | 1551 | 20,52 | 3,61   | 49,49 | 0       | 419,75  |
| STORAGE (àrea d'emmagatzematge)               | 204 | 1071 | 20,46 | 3,14   | 57,33 | 0       | 456,53  |
| TRANSFER (operacions de càrrega i descàrrega) | 67  | 647  | 11,08 | 2,1    | 21,25 | 0,01    | 121,38  |
| *   | 19  | 166  | 2,971 | 2,408  | 2,879 | 0,137   | 12,558  |

Taula 11: Descriptiva de la variable **KILLED** vs **TIPUSACT**.



La Figura 37 mostra el logaritme neperià de la variable **KILLED** segons la causa general de l'accident. En els gràfics es detecten diversos outliers per tenir un elevat nombre de morts.

Figura 37: Boxplot múltiple del **LN(KILLED)** vs **CAUSAG**.

Dels accidents produïts per una errada humana, HUMAN, destaquen: AN 1098 i 1168 amb 2.000 i 561 morts respectivament.

Dels accidents produïts per una fallida mecànica, MECHANICAL, destaquen: AN 318 amb 50 morts, AN 534 amb 49 morts, AN 989 amb 153 morts, AN 1750 amb 60 morts, AN 2401 amb 133 morts i AN 5438 amb 41 morts.

I dels accidents produïts per una fallida deguda a una reacció violenta, VREACTION, destaca el següent accident pel seu elevat nombre de morts: AN 1098 amb 2.000 morts.

Tal i com s'observa a la Taula 12, els accidents amb més morts en mitjana i més variabilitat són els causats per una fallida mecànica. Li segueixen els accidents produïts per una causa humana. En canvi, els accidents amb menys morts i menys variabilitat són els causats per un impacte, seguits dels accidents produïts per una reacció violenta (en el nombre de morts).

| CAUSA GENERAL            | N   | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|--------------------------|-----|-------|-------|--------|--------|---------|---------|
| 1 (Externa)              | 245 | 371   | 7,02  | 0      | 34,34  | 0       | 500     |
| 2 (Humana)               | 492 | 371   | 10,48 | 1      | 96,82  | 0       | 2.000   |
| 3 (Impacte)              | 65  | 102   | 1,538 | 0      | 2,469  | 0       | 10      |
| 4 (Instrumentació)       | 594 | 513   | 2,335 | 0      | 11,168 | 0       | 153     |
| 5 (Mecànica)             | 170 | 105   | 14,6  | 0      | 153,4  | 0       | 2.000   |
| 6 (Condicions<br>Procés) | 38  | 32    | 2,079 | 0      | 3,701  | 0       | 14      |
| 7 (Reacció Violenta)     | 28  | 3     | 1,964 | 0      | 4,542  | 0       | 20      |
| 8 (Serveis)              | 25  | 28    | 3,36  | 1      | 7,91   | 0       | 38      |
| *                        | 868 | 1.695 | 7,97  | 1      | 61,34  | 0       | 1.377   |

Taula 12: Descriptiva de la variable **KILLED** vs **CAUSAG**.

#### 4.4.6. Relació de Killed amb la quantitat de substància involucrada

A la Figura 38 s'hi pot observar el logaritme neperià de la variable **KILLED** segons el logaritme neperià de la quantitat de substància involucrada en l'accident en tones. No hi ha cap relació entre el nombre de morts i la quantitat de substància que ha intervingut en l'accident, ja que hi ha molta dispersió. S'observa que quan no hi ha cap mort, pot haver intervingut en l'accident qualsevol quantitat de substància, el mateix passa quan hi ha un mort, però no tant accentuat.

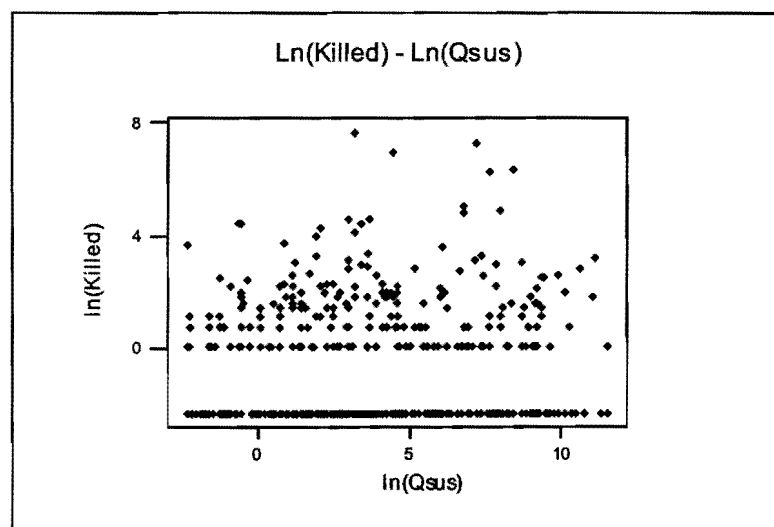
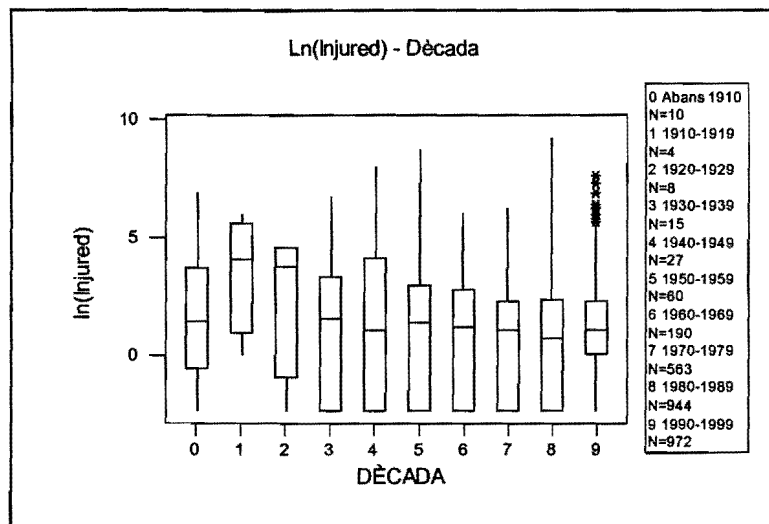


Figura 38: Boxplot múltiple del LN(KILLED) vs LN(QSUS).

## 4.5. Anàlisi bivariant per Injured

### 4.5.1. Relació de Injured amb la data de l'accident

A la Figura 39 s'hi pot observar el logaritme neperià de la variable **INJURED** segons la dècada en què es va produir l'accident. En general en totes les dècades hi ha molta dispersió, però la que menys té és la dècada dels 90. I en general el nombre



de ferits segueix una distribució no simètrica. A partir de la dècada dels 50, també es pot incloure la dècada dels 30, i fins els 80 el nombre de ferits s'assembla bastant tot i hi ha una petita disminució.

Figura 39: Boxplot múltiple del LN(INJURED) vs DÈCADA.

A la dècada dels noranta destaquen unes quantes observacions per tenir un nombre elevat de ferits:

AN 4030 amb 400 ferits  
 AN 4170 amb 540 ferits  
 AN 4714 amb 329 ferits  
 AN 4762 amb 300 ferits  
 AN 5065 amb 600 ferits  
 AN 5438 amb 403 ferits

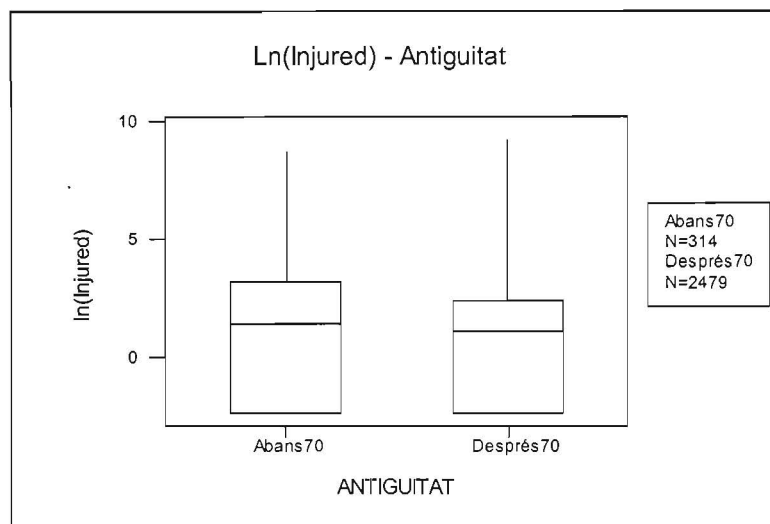
AN 6324 amb 2.000 ferits  
 AN 7186 amb 500 ferits  
 AN 7734 amb 330 ferits  
 AN 7861 amb 400 ferits  
 AN 8418 amb 1.500 ferits  
 AN 8603 amb 1.000 ferits.

Tal i com s'observa a la Taula 13, els accidents que tenen més ferits en mitjana són els que s'han produït a la dècada dels 40, seguits pels de la primera dècada. Els que tenen més variabilitat són els accidents de la dècada dels 50, seguits pels de la dècada dels 40. En canvi, els accidents que tenen menys ferits en mitjana i menys variabilitat són els que s'han produït en la dècada dels 70.

| Dècada | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|--------|-----|-------|-------|--------|-------|---------|---------|
| 0      | 10  | 5     | 116,4 | 4,5    | 312   | 0       | 1.000   |
| 1      | 4   | 2     | 133   | 65,5   | 181,8 | 1       | 400     |
| 2      | 8   | 3     | 52,3  | 44     | 43,1  | 0       | 100     |
| 3      | 15  | 5     | 72,2  | 5      | 206,2 | 0       | 804     |
| 4      | 27  | 23    | 158   | 3      | 577   | 0       | 3.000   |
| 5      | 60  | 48    | 121   | 4      | 773,7 | 0       | 6.000   |
| 6      | 190 | 99    | 22,32 | 3,5    | 52,62 | 0       | 434     |
| 7      | 563 | 256   | 13,68 | 3      | 38,76 | 0       | 500     |
| 8      | 944 | 375   | 41,5  | 2      | 473,6 | 0       | 9.999   |
| 9      | 972 | 1.558 | 19,55 | 3      | 97,34 | 0       | 2.000   |

Taula 13: Descriptiva de la variable **INJURED** vs **DÈCADA**.

A la Figura 40 està representat el logaritme neperià de la variable **INJURED** segons s'hagi produït l'accident abans o després de la dècada dels setanta. El



nombre de ferits té una distribució no simètrica. El nombre de ferits disminueix lleugerament a partir de la dècada dels 70.

Figura 40, Boxplot múltiple del **LN(INJURED)** vs **ANTIGUITAT**.

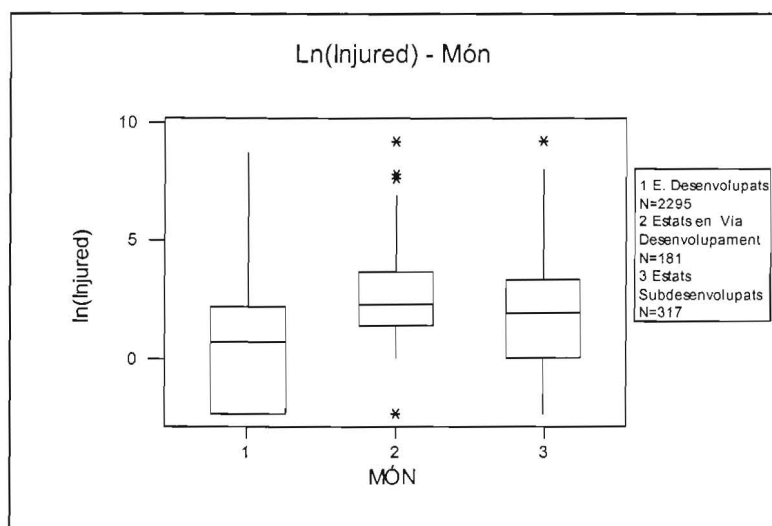
La Taula 14 resumeix el nombre de ferits en funció de la dècada de l'accident. Els accidents que es van produir abans de la dècada dels setanta tenen més ferits en mitjana i més variabilitat que els accidents que es van produir a partir de la dècada dels setanta, com ja s'havia observat en el gràfic anterior.

| Antiguitat      | N    | N*   | Mean  | Median | StDev  | Minimum | Maximum |
|-----------------|------|------|-------|--------|--------|---------|---------|
| 0 (Abans 70)    | 314  | 185  | 60,4  | 4      | 387    | 0       | 6000    |
| 1 (A partir 70) | 2479 | 2189 | 26,58 | 3      | 299,25 | 0       | 9999    |

Taula 14: Descriptiva de la variable **INJURED** vs **ANTIGUITAT**.

#### 4.5.2. Relació de Injured amb la situació geogràfica

A la Figura 41 està representat el logaritme neperià de la variable **INJURED** en funció del grau de desenvolupament de l'estat on es va produir l'accident. El



nombre de ferits més baix es troba en els estats desenvolupats, li segueix els estats subdesenvolupats. La distribució del nombre de ferits no és simètrica.

Figura 41: Boxplot múltiple del **LN(INJURED)** vs **MÓN**.



Dels accidents que es van produir en països en via de desenvolupament destaquen per tenir zero ferits, els següents: **AN**: 333, 345, 500, 916, 1038, 1374, 2167, 2959, 3109, 3585, 3814, 6731, 7087, 7188, 7450, 8251 i 8908. També destaquen per tenir molts ferits: **AN** 420 amb 2.500 ferits, **AN** 2072 amb 2.000 ferits i **AN** 2744 amb 9.999 ferits. I dels accidents que es van produir en estats subdesenvolupats destaca l'accident amb **AN** 1098 amb 9.999 ferits. Aquest és l'accident de Bhopal que va ser molt greu, ja que a més a més, hi van haver 2.000 víctimes mortals.

Tal i com s'observa a la Taula 15, els accidents que tenen més ferits en mitjana i més dispersió són els que es produeixen en països en via de desenvolupament. En canvi, els accidents amb menys ferits en mitjana i menys variabilitat són aquells que s'han produït en estats desenvolupats, tal i com s'observava abans en el gràfic.

| Món                              | N     | N*    | Mean  | Median | StDev | Minimum | Maximum |
|----------------------------------|-------|-------|-------|--------|-------|---------|---------|
| 1(Estats Desenvolupats)          | 2.295 | 1.950 | 15,84 | 2      | 138   | 0       | 6.000   |
| 2 (E. en via de Desenvolupament) | 181   | 135   | 123,3 | 10     | 781,5 | 0       | 9.999   |
| 3 (E. Subdesenvolupats)          | 317   | 289   | 82,6  | 7      | 596,8 | 0       | 9.999   |

Taula 15: Descriptiva de la variable **INJURED** vs **MÓN**.

### 4.5.3. Relació de Injured amb les substàncies involucrades

A la Figura 42 s'hi pot observar el logaritme neperià de la variable **INJURED** en funció de l'estat físic de les substàncies que han intervingut en l'accident. El nombre de ferits en accidents on intervenen substàncies en estat físic sòlid i pols, **DUST**, tenen una distribució semblant i els accidents on intervenen substàncies en estat físic líquid o gas pressuritzat, **PLGAS**, també.

En els accidents on han intervingut substàncies en estat físic sòlid destaca per tenir un nombre alt de ferits l'accident amb AN 2091 que té 3.000 ferits. En els accidents on han intervingut substàncies en estat físic líquid destaca per tenir un nombre alt de ferits els accidents amb AN 810 que té 6.000 ferits, AN 1098 amb 9.999 ferits i AN 2744 amb 9.999 ferits. I per últim en els accidents on han intervingut substàncies en estat físic gas destaca per tenir un nombre alt de ferits l'accident amb AN 8418 que té 1.500 ferits i AN 8603 que té 1.000 ferits.

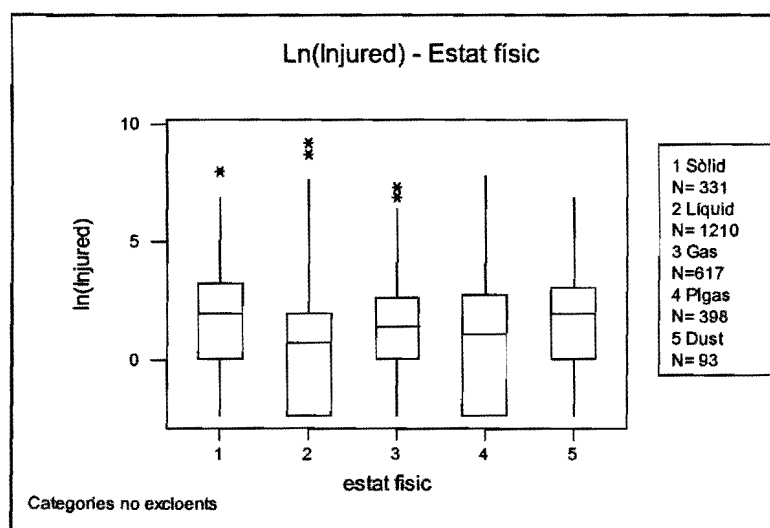


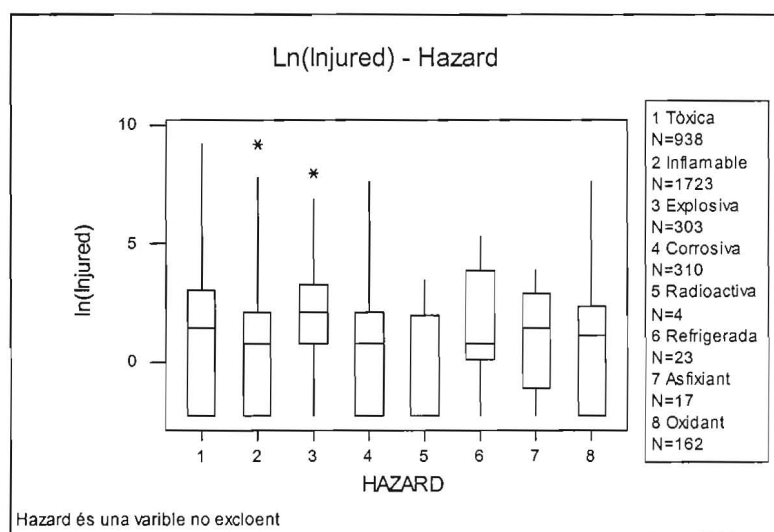
Figura 42: Boxplot múltiple del LN(INJURED) vs ESTAT\_FÍSIC.

Tal i com s'observa a la Taula 16, els accidents que tenen més ferits en mitjana són aquells en els que intervenen substàncies en estat físic sòlid. Li segueixen els accidents on han participat substàncies en estat físic líquid. Però, aquests últims tenen més variabilitat en el nombre de ferits. Per contra, els accidents que tenen menys ferits en mitjana i menys variabilitat són aquells en els que intervenen substàncies en estat físic gas (sinó es té en compte els missings).

| estat físic          | N     | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|----------------------|-------|-------|-------|--------|--------|---------|---------|
| 1 (Sòlid)            | 331   | 176   | 42,6  | 7      | 187,2  | 0       | 3.000   |
| 2 (Líquid)           | 1.210 | 1.318 | 34,2  | 2      | 446,1  | 0       | 9.999   |
| 3 (Gas)              | 617   | 360   | 25,27 | 4      | 94,81  | 0       | 1.500   |
| 4 (Gas Pressuritzat) | 398   | 173   | 30,9  | 3      | 167,35 | 0       | 2.500   |
| 5 (Pols)             | 93    | 43    | 30,1  | 7      | 108,7  | 0       | 1.000   |
| *                    | 250   | 349   | 9,52  | 3      | 20,4   | 0       | 236     |

Taula 16: Descriptiva de la variable **INJURED** vs **ESTAT\_FÍSIC**.

La Figura 43 mostra el logaritme neperià de la variable **INJURED** segons el risc de la substància que ha intervingut en els accidents. El nombre de ferits és diferent



segons el risc de la substància o substàncies que hagin participat en l'accident. Quan les substàncies són explosives la dispersió és menor.

Figura 43: Boxplot múltiple del **LN(INJURED)** vs **HAZARD**.

Dels accidents on ha intervingut substàncies inflamables destaca l'accident amb AN 1098 que va causar 9.999 ferits, pel seu gran nombre de ferits. I dels accidents on ha intervingut substàncies explosives destaca l'accident amb AN 2091 que va causar 3.000 ferits, pel mateix motiu.

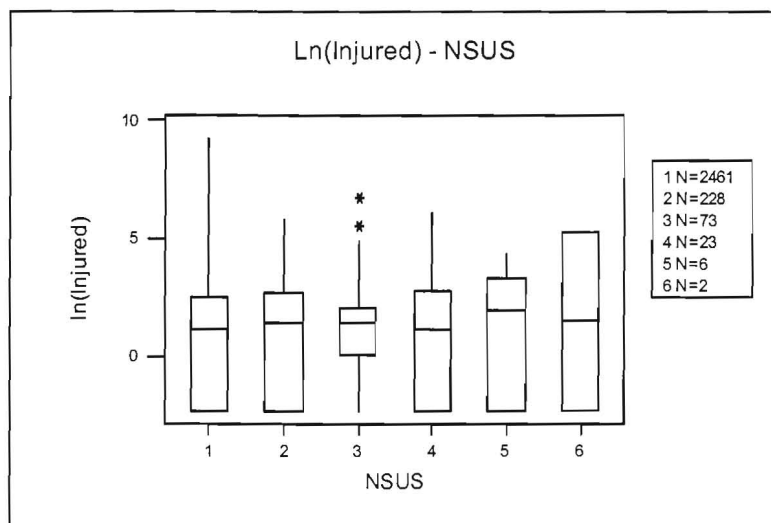
Tal i com s'observa a la Taula 17, els accidents amb més ferits en mitjana i més variabilitat són aquells on han participat substàncies tòxiques, seguits dels accidents on han participat substàncies explosives (en nombre de ferits).

D'altra banda, les substàncies radioactives tenen el nombre de ferits més baix i menys variabilitat, però aquesta substància té poques dades i no són representatives de la base de dades MHIDAS, li segueixen les substàncies asfixiants.

| HAZARD          | N     | N*    | Mean  | Median | StDev  | SEMean | Minimum | Maximum |
|-----------------|-------|-------|-------|--------|--------|--------|---------|---------|
| 1 (Tòxica)      | 938   | 583   | 55,3  | 4      | 509,1  | 16,6   | 0       | 9.999   |
| 2 (Inflamable)  | 1.723 | 1.546 | 19,91 | 2      | 253,23 | 6,1    | 0       | 9.999   |
| 3 (Explosiva)   | 303   | 141   | 46    | 8      | 201,5  | 11,6   | 0       | 3.000   |
| 4 (Corrosiva)   | 310   | 289   | 18,72 | 2      | 124,49 | 7,07   | 0       | 2.000   |
| 5 (Radioactiva) | 4     | 1     | 7,5   | 0      | 15     | 7,5    | 0       | 30      |
| 6 (Refrigerada) | 23    | 24    | 26,83 | 2      | 47,37  | 9,88   | 0       | 200     |
| 7 (Asfixiant)   | 17    | 15    | 11,88 | 4      | 15,3   | 3,71   | 0       | 48      |
| 8 (Oxidant)     | 162   | 145   | 33,2  | 3      | 174,1  | 13,7   | 0       | 2.000   |

Taula 17: Descriptiva de la variable **INJURED** vs **HAZARD**.

A la Figura 44 es pot observar el logaritme neperià de la variable **INJURED** segons el nombre de substàncies que han intervingut en l'accident. A partir de 3



substàncies el nombre de ferits s'incrementa en augmentar el nombre de substàncies. Els accidents on han participat 3 substàncies tenen menor variabilitat.

Figura 44: Boxplot múltiple del **LN(INJURED)** vs **NSUS**.

Dels accidents on han intervingut 3 substàncies destaquen els següents: **AN 1931** amb 804 ferits i **AN 2640** amb 250 ferits.

Tal i com s'observa a la Taula 18, els accidents on hi ha hagut més ferits en mitjana han estat aquells on han participat sis substàncies, encara que només hi ha dos accidents amb aquestes característiques. Li segueixen els accidents on han participat quatre substàncies. Cal destacar els accidents on només intervé una substància per tenir la dispersió més gran, però també és cert que són els accidents més nombrosos. En canvi, els que han tingut menys ferits en mitjana han estat els accidents on han intervingut dues substàncies. I els que tenen menys variabilitat són aquells on han participat cinc substàncies, però només té sis observacions, li segueixen els accidents on han intervingut dues substàncies.

| NSUS | N     | N*    | Mean  | Median | StDev  | Minimum | Maximum |
|------|-------|-------|-------|--------|--------|---------|---------|
| 1    | 2.461 | 2.170 | 31,69 | 3      | 329,91 | 0       | 9.999   |
| 2    | 228   | 146   | 17,21 | 4      | 39,18  | 0       | 350     |
| 3    | 73    | 32    | 23,8  | 4      | 99,2   | 0       | 804     |
| 4    | 23    | 23    | 39,3  | 3      | 98,8   | 0       | 439     |
| 5    | 6     | 2     | 19,2  | 7,5    | 30,7   | 0       | 80      |
| 6    | 2     | 1     | 100   | 100    | 141    | 0       | 200     |

Taula 18: Descriptiva de la variable **INJURED** vs **NSUS**.

#### 4.5.4. Relació de Injured amb el tipus d'accident

A la Figura 45 es mostra el logaritme neperià de la variable **INJURED** segons el tipus d'accident. El tipus d'accident que té menys variabilitat en el nombre de ferits és el produït per una explosió. Els accidents que es produeixen per un incendi, **FIRE**, tenen una distribució semblant als accidents que es produeixen per una fuga, **RELEASE**.

Dels accidents que es van produir per una explosió destaquen pel seu elevat nombre de ferits: **AN 420** amb 2.500 ferits, **AN 1931** amb 804 ferits, **AN 2006** amb 1.000 ferits i **AN 2091** amb 3.000 ferits.

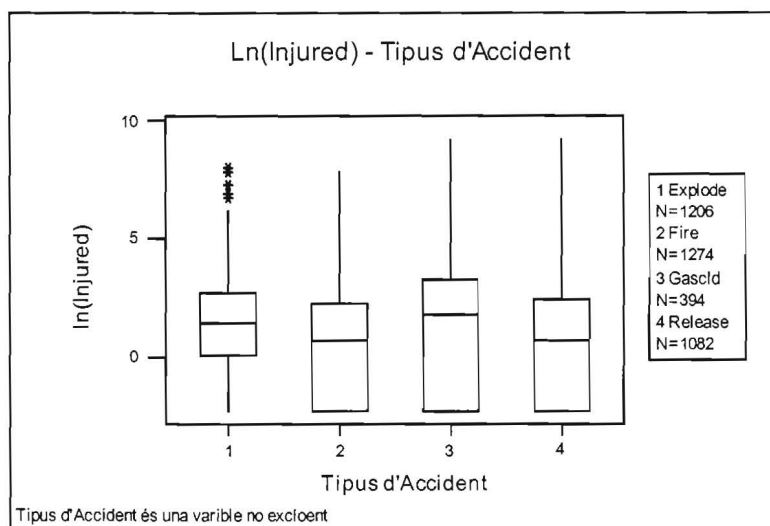


Figura 45: Boxplot múltiple del **LN(INJURED)** vs **TIPUSACC**.

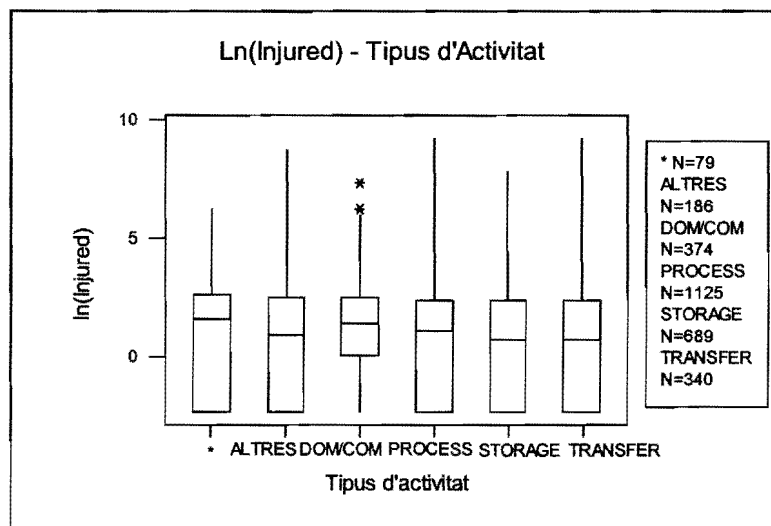
Tal i com s'observa a la Taula 19, els accidents que provoquen més ferits en mitjana i més variabilitat són els que es produeixen per un núvol de gas. Li segueixen els que es produeixen per una fuga. D'altra banda, els accidents amb menys ferits i menys variabilitat són els accidents produïts per un incendi (si no es té en compte els missings).

| Tipus Accident   | N    | N*   | Mean  | Median | StDev  | Minimum | Maximum |
|------------------|------|------|-------|--------|--------|---------|---------|
| 1 (Explosió)     | 1206 | 469  | 26,16 | 4      | 137,83 | 0       | 3000    |
| 2 (Incendi)      | 1274 | 966  | 16,85 | 2      | 96,33  | 0       | 2500    |
| 3 (Núvol de gas) | 394  | 122  | 60,4  | 6      | 516,7  | 0       | 9999    |
| 4 (Fuga)         | 1082 | 1184 | 45,9  | 2      | 477,1  | 0       | 9999    |
| *                | 79   | 126  | 15,22 | 42,48  | 4,78   | 0       | 300     |

Taula 19: Descriptiva de la variable **INJURED** vs **TIPUSACC**.

### 4.5.5. Relació de Injured amb el tipus d'activitat

A la Figura 46 es pot observar el logaritme neperià de la variable **INJURED** en funció del tipus d'activitat que es duia a terme en el moment de l'accident. El



nombre de ferits segons el tipus d'activitat que es duia a terme en el moment de l'accident segueix una distribució asimètrica.

Figura 46: Boxplot múltiple del LN(INJURED) vs TIPUSACT.

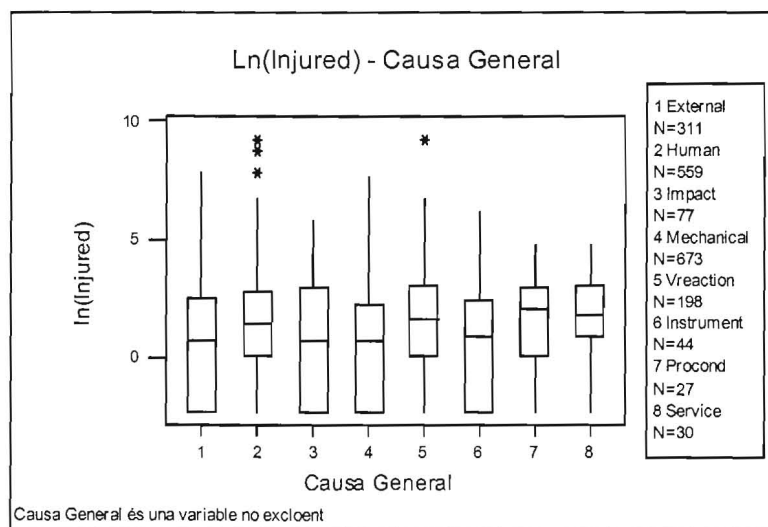
D'entre els accident que es van produir durant activitats domèstiques o comercials, DOM/COM, destaquen per tenir un gran nombre de ferits els següents accidents: AN 7186 amb 500 ferits i AN 8418 amb 1.500 ferits.

Tal i com s'observa a la Taula 20, el tipus d'activitat amb més ferits en mitjana i més desviació tipus és la categoria TRANSFER, durant operacions de càrrega i descàrrega. Li segueixen els de la categoria ALTRES, tasques auxiliars de la indústria. Per contra, l'activitat amb menys ferits i menys variabilitat és DOM/COM, en locals domèstics o comercials, li segueix STORAGE, en plantes d'emmagatzematge.

| Tipus Activitat                               | N     | N*  | Mean  | Median | StDev  | Minimum | Maximum |
|---|-------|-----|-------|--------|--------|---------|---------|
| ALTRES (tasques auxiliars de la indústria)    | 186   | 138 | 51    | 2,5    | 445,5  | 0       | 6.000   |
| DOM/COM (locals domèstics o comercials)       | 374   | 415 | 18,2  | 4      | 86,27  | 0       | 1.500   |
| PROCESS (plantes de procés)                   | 1.125 | 755 | 26,78 | 3      | 303,99 | 0       | 9.999   |
| STORAGE (plantes d'emmagatzematge)            | 689   | 586 | 23,42 | 2      | 133,44 | 0       | 2.500   |
| TRANSFER (operacions de càrrega i descàrrega) | 340   | 374 | 59,5  | 2      | 576,2  | 0       | 9.999   |
| *   | 79    | 106 | 26,2  | 5      | 76,76  | 0       | 500     |

Taula 20: Descriptiva de la variable **INJURED** vs **TIPUSACT**.

Tal i com s'observa a la Figura 47 el nombre de ferits varia en funció de quina



ha sigut la causa de l'accident. Segueix una distribució no simètrica. La causa de l'accident amb menys variabilitat és una fallida en serveis, SERVICE.

Figura 47: Boxplot múltiple del **LN(INJURED)** vs **CAUSAG**.

Dels accident produïts per una errada humana destaquen pel seu gran nombre de ferits: **AN 420** amb 2.500 ferits, **AN 810** amb 6.000 ferits, **AN 1098** amb 9.999 ferits i **AN 2744** amb 9.999 ferits

I dels accident produïts per una reacció violenta destaca el següent: **AN 1098** amb 9.999 ferits.



Tal i com s'observa a la Taula 21, la causa que provoca més ferits en mitjana i més variabilitat és la originada per una reacció violenta, VREACTION, li segueix els accidents produïts per una errada humana, HUMAN. Per contra, la causa que provoca menys ferits i menys variabilitat és l'originada per una variació de les condicions de procés, PROCOND, seguida dels accidents produïts per un impacte, IMPACT, en el nombre de ferits i els accidents produïts per una fallida de serveis, SERVICE, en variabilitat.

| Causa General         | N    | N*   | Mean  | Median | Stdev  | SEMean | Minimum | Maximum |
|-----------------------|------|------|-------|--------|--------|--------|---------|---------|
| 1 (Externa)           | 311  | 305  | 30,27 | 2      | 161,08 | 9,13   | 0       | 2500    |
| 2 (Humana)            | 559  | 304  | 73,8  | 4      | 658,3  | 27,8   | 0       | 9999    |
| 3 (Impacte)           | 77   | 90   | 18,87 | 2      | 45,47  | 5,18   | 0       | 329     |
| 4 (Mecànica)          | 673  | 434  | 20,68 | 2      | 116,24 | 4,48   | 0       | 2000    |
| 5 (Reacció violenta)  | 198  | 77   | 74,6  | 5      | 712,2  | 50,6   | 0       | 9999    |
| 6 (Instrumentació)    | 44   | 26   | 23,2  | 2,5    | 77,3   | 11,7   | 0       | 500     |
| 7 (Condicions Procés) | 27   | 4    | 18,04 | 8      | 29,7   | 5,71   | 0       | 120     |
| 8 (Serveis)           | 30   | 23   | 19,3  | 6      | 31,62  | 5,77   | 0       | 123     |
| *                     | 1282 | 1281 | 19,27 | 3      | 109,95 | 3,07   | 0       | 3000    |

Taula 21: Descriptiva de la variable **INJURED** vs **CAUSAG**.

#### 4.5.6. Relació de Injured amb la quantitat de substància involucrada

El logaritme neperià de la variable **INJURED**, veure Figura 48, no té cap relació amb la quantitat de substància involucrada en l'accident. Hi ha molta dispersió. S'observa que quan no hi ha cap ferit, pot haver intervingut en l'accident qualsevol quantitat de substància i el mateix passa quan hi ha un ferit, però no tant clar.

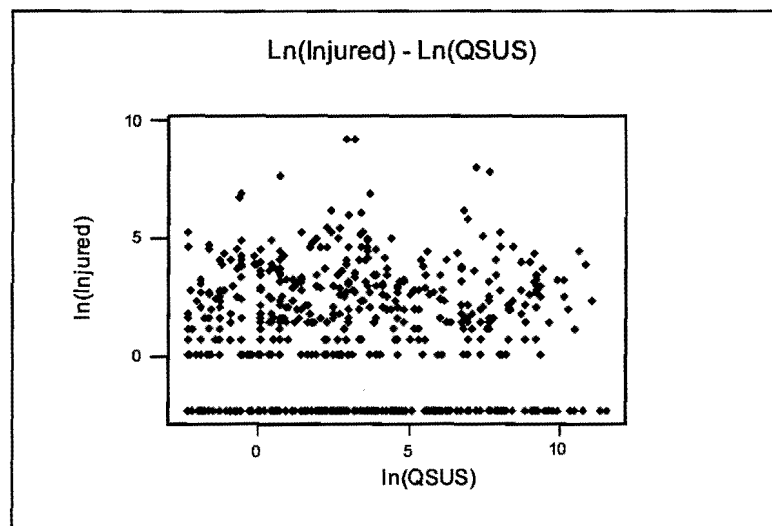


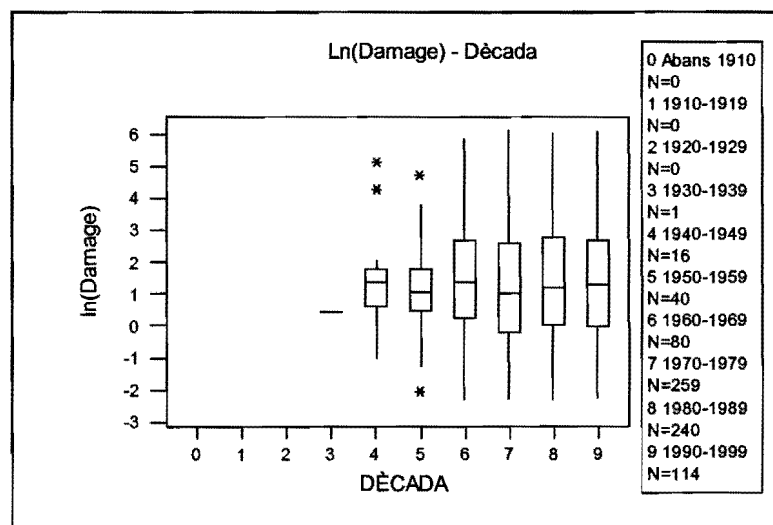
Figura 48: Gràfic del LN(INJURED) vs LN(QSUS).

## 4.6. Anàlisi bivariant per Damage

### 4.6.1. Relació de Damage amb la data de l'accident

La Figura 49 conté la representació del logaritme neperià de l'actualització econòmica del dany material segons la dècada en la que es va produir l'accident. Pels accidents anteriors a l'any 1.913 no es pot calcular l'actualització econòmica del dany material, ja que per aquests anys no hi ha índex d'actualització del cost Marshall & Stevens. A partir de la dècada dels 60 els boxplots s'assemblen bastant,

el que significa que



l'estimació econòmica del dany material és semblant en les diferents dècades. Tot i que hi ha molta dispersió, en comparació amb la dècada dels 40 i 50.

Figura 49: Boxplot múltiple del LN(DAMAGE) vs DÈCADA.

A la dècada dels quaranta hi ha alguns accidents que destaquen per tenir una estimació econòmica del dany material molt elevada, aquests són: AN 1755 amb uns danys de 71,048 milions de dòlars i AN 1818 amb uns danys de 163,945 milions de dòlars.

A la dècada dels cinquanta destaca per tenir una estimació econòmica del dany material inferior, en comparació a la resta d'accidents d'aquesta dècada, l'accident amb AN 1978 que té uns danys de 0,029 milions de dòlars. I també destaca l'accident amb AN 1693 que té uns danys de 112,267 milions de dòlars, per tenir una estimació econòmica del dany material elevada.

Tal i com s'observa a la Taula 22, els accidents que tenen un cost més elevat en mitjana i més variabilitat són els de la dècada dels 80, seguits pels succeïts a la dècada dels noranta. Per contra, els accidents amb una repercussió econòmica menor en mitjana i menys variabilitat són els ocorreguts a la dècada dels 50, si no es té en compte la dècada dels 30, ja que només hi ha un accident que té actualització econòmica.

| Dècada | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|--------|-----|-------|-------|--------|-------|---------|---------|
| 0      | 0   | 15    | *     | *      | *     | *       | *       |
| 1      | 0   | 6     | *     | *      | *     | *       | *       |
| 2      | 0   | 11    | *     | *      | *     | *       | *       |
| 3      | 1   | 19    | 1,434 | 1,434  | *     | 1,434   | 1,434   |
| 4      | 16  | 34    | 17,7  | 3,8    | 42,5  | 0,3     | 163,9   |
| 5      | 40  | 68    | 8,77  | 2,83   | 18,86 | 0,03    | 112,27  |
| 6      | 80  | 209   | 14,94 | 3,78   | 40,72 | 0       | 340,7   |
| 7      | 259 | 560   | 15,84 | 2,68   | 40,97 | 0       | 456,53  |
| 8      | 240 | 1.079 | 21,26 | 3,13   | 55,49 | 0       | 402,42  |
| 9      | 114 | 2.416 | 18,52 | 3,53   | 51,48 | 0,01    | 419,75  |

Taula 22: Descriptiva de la variable **DAMAGE** vs **DÈCADA**.

A la Figura 50 s'hi pot observar l'estimació econòmica del dany material segons s'hagi produït l'accident abans o després de la dècada dels setanta. Abans de la dècada dels 70 hi ha menys dispersió. Dels accidents produïts abans de la dècada dels setanta destaca per tenir una estimació econòmica del dany material molt eleva l'accident que té AN 1959 amb 340,704 milions de dòlars.

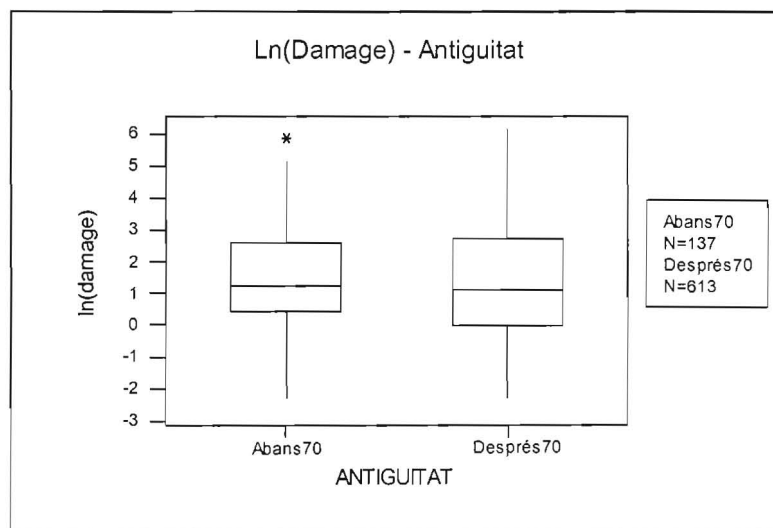


Figura 50: Boxplot múltiple del **LN(DAMAGE)** vs **ANTIGUITAT**.

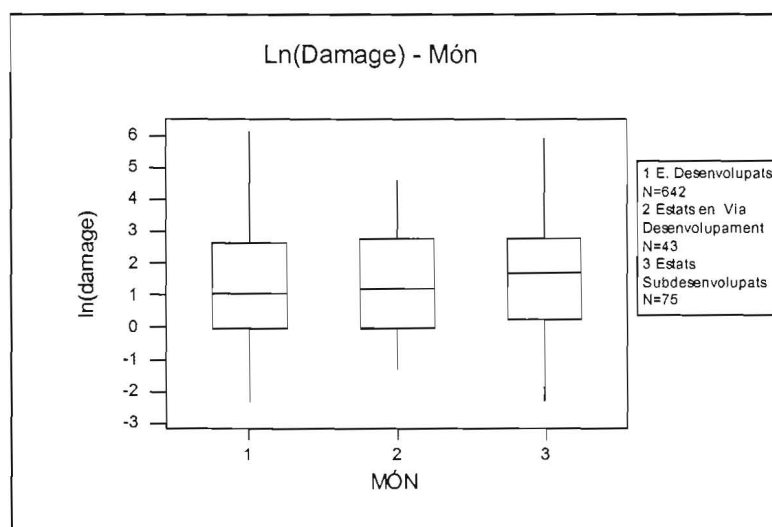
Tal i com s'observa a la Taula 23, els danys que han causat els accidents produïts després dels anys 70, aquest inclòs, són majors en mitjana i variabilitat, la qual cosa ja s'observava en el gràfic anterior.

| ANTIGUITAT      | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|-----------------|-----|-------|-------|--------|-------|---------|---------|
| 0 (Abans 70)    | 137 | 362   | 13,37 | 3,32   | 35,72 | 0       | 340,7   |
| 1 (A partir 70) | 613 | 4.055 | 18,46 | 2,92   | 49,04 | 0       | 456,53  |

Taula 23: Descriptiva de la variable **DAMAGE** vs **ANTIGUITAT**.

## 4.6.2. Relació de Damage amb la situació geogràfica

A la Figura 51 es mostra l'estimació econòmica del dany material en funció del



grau de desenvolupament de l'estat on es va produir l'accident. La dispersió és la semblant.

Figura 51: Boxplot múltiple del LN(DAMAGE) vs MÓN.

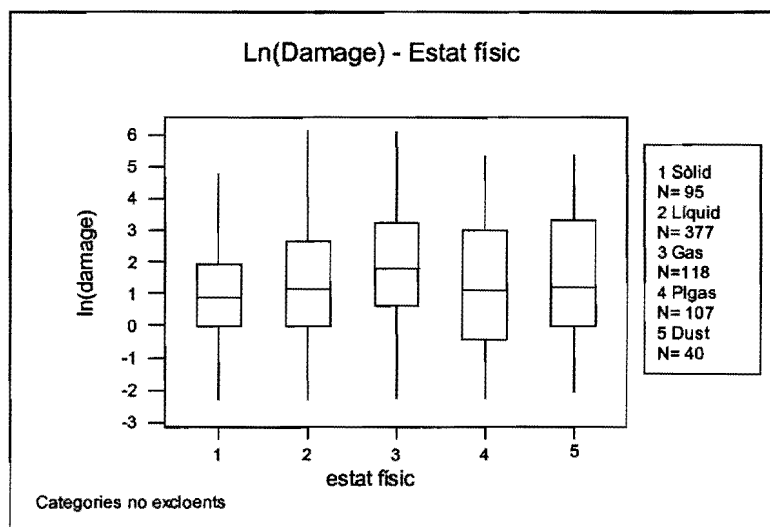
A la Taula 24, els accidents que tenen un cost més gran en mitjana i més variabilitat són els que es produeixen en països subdesenvolupats. I en canvi, els accidents amb un cost menor en mitjana i menys variabilitat són els que es produeixen en estats en via de desenvolupament.

| MÓN                              | N   | N*   | Mean  | Median | StDev | Minimum | Maximum |
|----------------------------------|-----|------|-------|--------|-------|---------|---------|
| 1 (Estats Desenvolupats)         | 632 | 3613 | 17,28 | 2,81   | 47,1  | 0       | 456,53  |
| 2 (E. en via de Desenvolupament) | 43  | 273  | 12,56 | 3,17   | 20,26 | 0,18    | 94,85   |
| 3 (E. Subdesenvolupats)          | 75  | 531  | 22,47 | 5,42   | 55,54 | 0       | 358,59  |

Taula 24: Descriptiva de la variable DAMAGE vs MÓN.

### 4.6.3. Relació de Damage amb les substàncies involucrades

La Figura 52 mostra el gràfic del logaritme neperià de la variable **DAMAGE** en funció de l'estat físic de les substàncies que han participat en l'accident. El dany econòmic dels accidents on intervenen substàncies en estat físic líquid i gas tenen una distribució similar, encara que el dany material és més elevat quan l'estat físic de la substància que intervé en l'accident és gas. Els accidents on intervenen



substàncies en estat físic gas pressuritzat, PLGAS, i pols, DUST, tenen una distribució semblant, encara que aquest últim té un dany material més elevat.

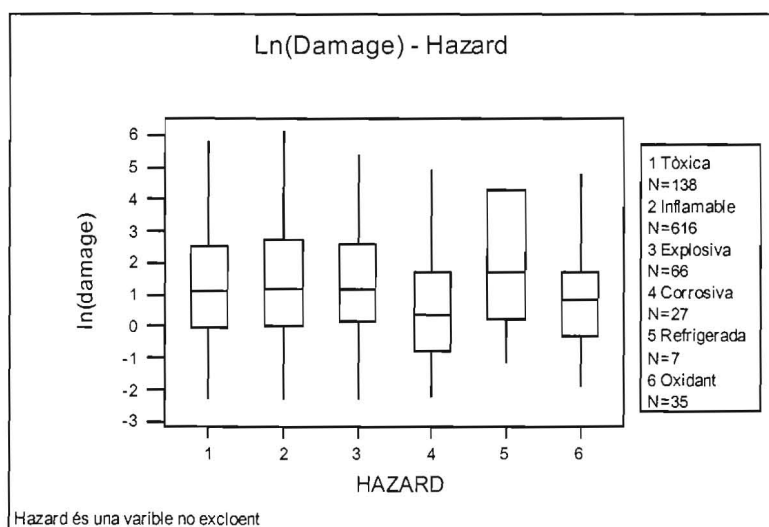
Figura 52: Boxplot múltiple del LN(DAMAGE) vs ESTAT\_FÍSIC.

A la Taula 25, s'observa com els accidents que tenen una estimació econòmica del dany material més gran en mitjana i més variabilitat són aquells on intervenen substàncies en estat físic gas. Li segueixen els accidents on han participat substàncies en estat físic pols, en mitjana. En canvi, els accidents que tenen una estimació econòmica menor en mitjana i menys variabilitat són aquells en els que intervenen substàncies en estat físic sòlid. Li segueixen en mitjana els accidents on participen substàncies en estat físic líquid en el cost.

| Estat_físic          | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|----------------------|-----|-------|-------|--------|-------|---------|---------|
| 1 (Sòlid)            | 95  | 412   | 8,5   | 2,27   | 17,93 | 0       | 114,97  |
| 2 (Liquid)           | 377 | 2.151 | 16,75 | 3,11   | 48,88 | 0       | 456,53  |
| 3 (Gas)              | 118 | 859   | 29,29 | 5,71   | 65,83 | 0,01    | 419,75  |
| 4 (Gas Pressuritzat) | 107 | 464   | 17,43 | 2,87   | 34,38 | 0,01    | 194,7   |
| 5 (Pols)             | 40  | 96    | 25,05 | 3,24   | 44,3  | 0,03    | 211,69  |

Taula 25: Descriptiva de la variable **DAMAGE** vs **ESTAT\_FÍSIC**.

A la Figura 53 s'hi pot observar el logaritme neperià de la variable **DAMAGE** en funció del risc de la substància o substàncies que han intervingut en l'accident. L'estimació econòmica del dany material en els accidents on han participat substàncies tòxiques, inflamables i explosives és semblant. En canvi en les



substàncies corrosives i oxidants l'estimació econòmica del dany material és més petita i en les substàncies refrigerades és més gran.

Figura 53: Boxplot múltiple del **LN(DAMAGE)** vs **HAZARD**.

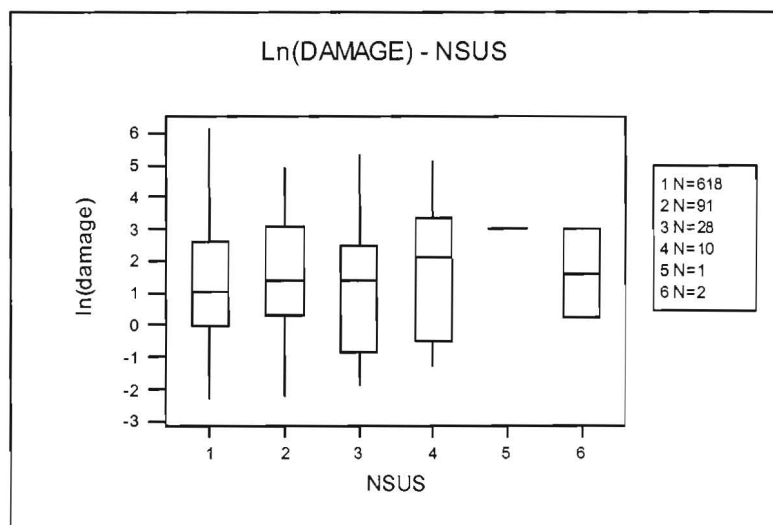
A la taula que es mostra a continuació s'observa que els accidents que tenen una estimació econòmica del dany material més elevat en mitjana són aquells on han intervingut substàncies refrigerades. I els que tenen més variabilitat en l'estimació econòmica del dany material són aquells on han participat substàncies inflamables. Per contra, els accidents que tenen una estimació econòmica del dany material més baixa en mitjana i menys variabilitat són aquells on han intervingut substàncies oxidants.



| HAZARD          | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|-----------------|-----|-------|-------|--------|-------|---------|---------|
| 1 (Tòxica)      | 138 | 1.383 | 11,65 | 2,96   | 30,63 | 0,01    | 322,73  |
| 2 (Inflamable)  | 616 | 2.653 | 18,69 | 3,24   | 49,96 | 0       | 456,53  |
| 3 (Explosiva)   | 66  | 378   | 18,17 | 3,17   | 42,39 | 0       | 211,69  |
| 4 (Corrosiva)   | 27  | 572   | 14,79 | 1,36   | 34,5  | 0,01    | 137,1   |
| 5 (Radioactiva) | 0   | 5     | *     | *      | *     | *       | *       |
| 6 (Refrigerada) | 7   | 40    | 23,4  | 5,4    | 32,6  | 0,2     | 71      |
| 7 (Asfixiant)   | 0   | 32    | *     | *      | *     | *       | *       |
| 8 (Oxidant)     | 35  | 272   | 9,88  | 2,15   | 24,38 | 0,06    | 114,97  |

Taula 26: Descriptiva de la variable **DAMAGE** vs **HAZARD**.

A la Figura 54 s'hi pot observar el logaritme neperià de l'estimació econòmica del dany material en funció del nombre de substàncies que han intervingut en l'accident. Quan en un accident intervé 1 ó 2 substàncies l'estimació econòmica del



dany material és semblant, però és més gran si intervenen 2 substàncies. El mateix passa per 3 i 4 substàncies.

Figura 54: Boxplot múltiple del **LN(DAMAGE)** vs **NSUS**.

Tal i com s'observa a la Taula 27, els accidents amb un cost més elevat en mitjana i més variabilitat són aquells on han participat quatre substàncies. Li segueix en cost elevat els accidents on han intervingut una substància, sense tenir en compte els accidents on han intervingut cinc substàncies, perquè només es coneix l'estimació econòmica d'un.

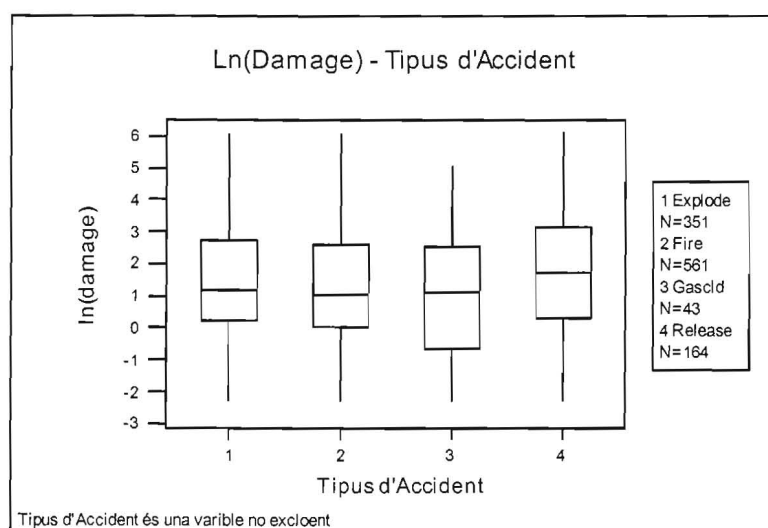
D'altra banda, els accidents on la repercussió econòmica ha estat menor i menys variabilitat són aquells on han intervingut sis substàncies, encara que només té dos accidents, seguits dels accidents on han participat dues substàncies.

| NSUS | N   | N*   | Mean   | Median | StDev | Minimum | Maximum |
|------|-----|------|--------|--------|-------|---------|---------|
| 1    | 618 | 4013 | 17,57  | 2,81   | 49,59 | 0       | 456,53  |
| 2    | 91  | 283  | 16,62  | 3,92   | 26,94 | 0,01    | 137,1   |
| 3    | 28  | 77   | 16,97  | 3,97   | 39,64 | 0,06    | 194,7   |
| 4    | 10  | 36   | 26,4   | 10,4   | 49,6  | 0,2     | 163,9   |
| 5    | 1   | 7    | 20,396 | 20,396 | *     | 20,396  | 20,396  |
| 6    | 2   | 1    | 10,31  | 10,31  | 12,95 | 1,15    | 19,47   |

Taula 27: Descriptiva de la variable **DAMAGE** vs **NSUS**.

#### 4.6.4. Relació de Damage amb el tipus d'accident

A la Figura 55 està representat el logaritme neperià de la variable **DAMAGE** en funció del tipus d'accident. L'estimació econòmica del dany material pels accidents



produïts per una explosió, EXPLODE, o un incendi, FIRE, és bastant semblant.

Figura 55: Boxplot múltiple del **LN(DAMAGE)** vs **TIPUSACC**.

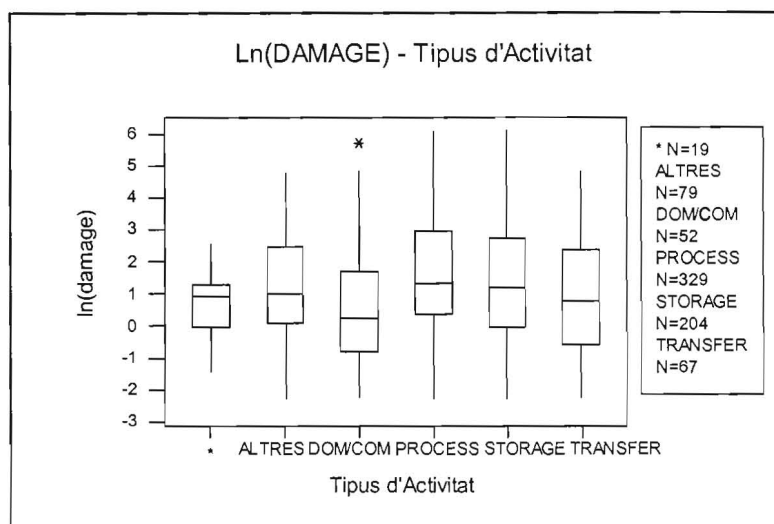
La Taula 28, mostra com els accidents que tenen una estimació econòmica del dany material més elevat en mitjana i més variabilitat són els que es produeixen per una fuga. En canvi, els accidents que tenen una estimació econòmica del dany material més baixa i menys variabilitat són els produïts per un núvol de gas, seguits pels accidents que es produeixen per un incendi.

| TIPUS D'ACCIDENT | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|------------------|-----|-------|-------|--------|-------|---------|---------|
| 1 (Explosió)     | 351 | 1.324 | 18,62 | 3,24   | 48,32 | 0,01    | 419,75  |
| 2 (Incendi)      | 561 | 1.679 | 15,6  | 2,85   | 44,59 | 0       | 419,75  |
| 3 (Núvol de Gas) | 43  | 473   | 14,31 | 2,92   | 30,44 | 0       | 155,76  |
| 4 (Fuga)         | 164 | 2.102 | 26,08 | 5,61   | 60,54 | 0,01    | 456,53  |

Taula 28: Descriptiva de la variable **DAMAGE** vs **TIPUSACC**.

#### 4.6.5. Relació de Damage amb el tipus d'activitat

El dany material és diferent segons el tipus d'activitat que es duia a terme en el moment de l'accident, tal i com s'observa a la Figura 56. L'estimació econòmica del dany material pels accidents produïts en locals domèstics o comercials, DOM/COM, en la planta de procés, PROCESS, i en tasques auxiliars a la indústria, ALTRES,



tenen una distribució semblant. El mateix passa amb els accidents produïts en operacions de càrrega i descàrrega, TRANSFER, i en plantes d'emmagatzematge.

Figura 56: Boxplot múltiple del **LN(DAMAGE)** vs **TIPUSACT**.

De tots els accidents produïts en locals domèstics o comercials destaca per tenir una estimació econòmica del dany material molt elevada el que té **AN 5151** amb uns danys de 303.173 milions de dòlars.

Tal i com s'observa a la Taula 29, els accidents que tenen una repercussió econòmica major, en mitjana, i més variabilitat són els que s'han originat en plantes de procés, **PROCESS**, i els que s'han produït en plantes d'emmagatzematge, **STORAGE**. En canvi, els accidents produïts durant tasques auxiliars de la indústria, **ALTRES**, són els que tenen una repercussió econòmica menor, en mitjana i menys variabilitat, si no es té en compte els missings.

| TIPUS D'ACTIVITAT                             | N   | N*    | Mean  | Median | StDev | Minimum | Maximum |
|---|-----|-------|-------|--------|-------|---------|---------|
| ALTRES (tasques auxiliars de la indústria)    | 79  | 245   | 9,73  | 2,71   | 18,18 | 0       | 114,97  |
| DOM/COM (locals domèstics o comercials)       | 52  | 737   | 12,59 | 1,16   | 45,28 | 0,01    | 303,17  |
| PROCESS (planta de procés)                    | 329 | 1551  | 20,52 | 3,61   | 49,49 | 0       | 419,75  |
| STORAGE (planta d'emmagatzematge)             | 204 | 1.071 | 20,46 | 3,14   | 57,33 | 0       | 456,53  |
| TRANSFER (operacions de càrrega i descàrrega) | 67  | 647   | 11,08 | 2,1    | 21,25 | 0,01    | 121,38  |
| *   | 19  | 166   | 2,971 | 2,408  | 2,879 | 0,137   | 12,558  |

Taula 29: Descriptiva de la variable **DAMAGE** vs **TIPUSACT**.

La Figura 57 mostra el logaritme neperià de l'estimació econòmica del dany material actualitzada amb l'índex Marshall & Stevens segons la causa general de l'accident. S'observa com el cost és diferents segons la causa que ha provocat l'accident. Quan l'accident es produït per una variació en les condicions de procés la variabilitat és menor.

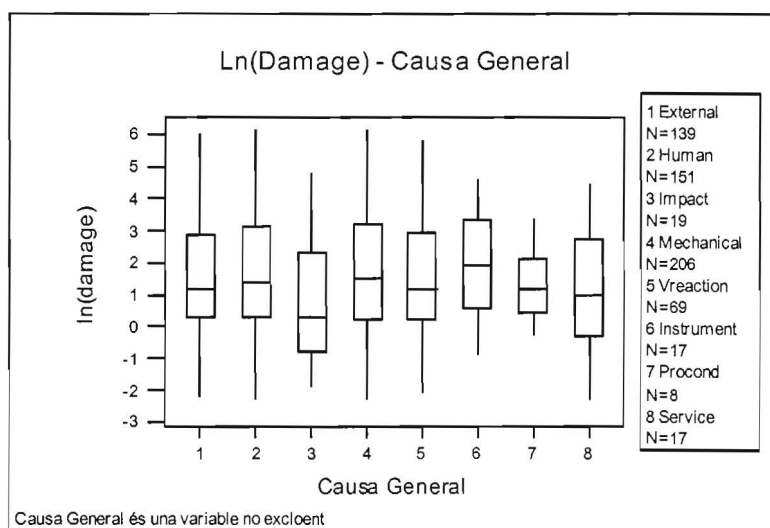


Figura 57: Boxplot múltiple del LN(DAMAGE) vs CAUSAG.

Segons la Taula 30 els accidents que tenen una estimació econòmica del dany material major en mitjana i més variabilitat són aquells que tenen com a causa general una errada humana. Li segueixen els que es produeixen per una fallida d'instruments. Per contra, els que tenen una estimació inferior i menys variabilitat són els accidents produïts per una reacció violenta, però només hi ha vuit accidents. Li segueixen els accidents produïts per una fallida de serveis.

| CAUSA GENERAL         | N   | N*  | Mean  | Median | StDev | Minimum | Maximum |
|-----------------------|-----|-----|-------|--------|-------|---------|---------|
| 1 (Externa)           | 139 | 477 | 19,97 | 3,24   | 51,75 | 0,01    | 402,42  |
| 2 (Humana)            | 151 | 712 | 25,94 | 3,92   | 62,47 | 0,01    | 456,53  |
| 3 (Impacte)           | 19  | 148 | 12,12 | 1,26   | 28,95 | 0,06    | 114,97  |
| 4 (Instrument)        | 206 | 901 | 23,56 | 4,58   | 52,47 | 0,01    | 456,53  |
| 5 (Mecànica)          | 69  | 206 | 20,13 | 3,24   | 48,94 | 0,03    | 322,73  |
| 6 (Condicions Procés) | 17  | 53  | 22,66 | 6,66   | 30,73 | 0,33    | 96,32   |
| 7 (Reacció Violenta)  | 8   | 23  | 6,73  | 3,27   | 8,86  | 0,71    | 27,49   |
| 8 (Serveis)           | 17  | 36  | 11,56 | 2,51   | 21,02 | 0       | 85,48   |

Taula 30: Descriptiva de la variable DAMAGE vs CAUSAG.

#### 4.6.6. Relació de Damage amb la quantitat de substància involucrada

A la Figura 58 es pot observar l'estimació econòmica del dany material en funció del logaritme neperià de la quantitat de substància que ha intervingut en l'accident. No hi ha cap relació entre l'actualització econòmica del dany material i la quantitat de substància que ha intervingut en l'accident. Cal destacar que hi ha molta dispersió.

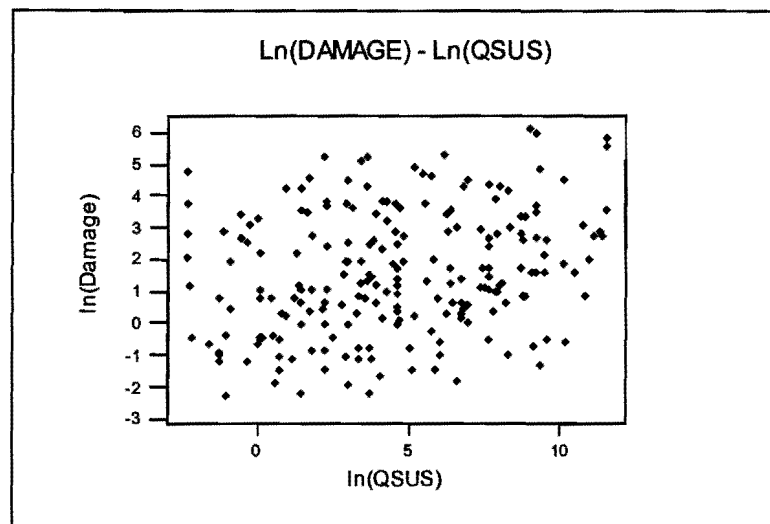


Figura 58: Boxplot múltiple del LN(DAMAGE) vs LN(QSUS).

## **4.7. Conclusions**

Les variables resposta, **KILLED**, **INJURED**, **EVACUATED** i **DAMAGE** s'han transformat fent el logaritme neperià de les dades més 0,1, ja que les tres primeres són comptetjos i prendre logaritmes acostuma a linealitzar el model i estabilitzar la variància. També s'ha transformat la variable **DAMAGE** per normalitzar-la i poder linealitzar el model. Però, s'ha descartat com a variable resposta **EVACUATED**, ja que la transformació logarítmica no ha solucionat la no normalitat, té molts zeros i uns que continuen destacant molt sobre la resta. La variable **QSUS** també s'ha transformat prenent logaritmes perquè té una distribució molt asimètrica.

Els accidents produïts a partir de la dècada dels 70 fins l'any 98, representen el 90,3% del total. Això és degut a que les dades es van començar a recollir a principis dels anys 80, i a que el nombre de fàbriques ha augmentat molt en aquests últims anys. El 82,9% s'han produït en països desenvolupats.

El 52,6% dels accidents són amb substàncies en estat físic líquid. Li segueix l'estat físic gas amb un 20,7%. El risc més freqüent de les substàncies que intervenen en els accidents és inflamable, amb un 52,5% dels accidents, i seguit del risc tòxic, 24,4%. El més freqüent és que en un accident hi intervingui una única substància, ja que aquests representen un 89,6% del total d'accidents.

El tipus d'accident més freqüent és el que es produeix per fuga, que representa el 29,9%. Li segueixen els accidents produïts per un incendi, amb 19,9%, i per últim els accidents produïts per una explosió que posteriorment esdevenen en un incendi, un 13,7%. Els accidents més freqüents s'han produït durant tasques en unitats o àrees d'una planta de procés, un 37,7%. Li segueixen els accidents produïts en tasques auxiliars de la indústria, 25,6%. Les causes generals que han provocat més accidents són les degudes a una errada mecànica, 34,8% i a una errada humana, 27,1%.

Pel que fa al nombre de morts, els accidents que tenen un major nombre de morts són els produïts a la dècada dels anys 20, i en general d'abans de la dècada dels 70, en els països subdesenvolupats. El nombre de morts també és alt quan les substàncies que intervenen estan en estat físic pols i sòlid, quan són explosives i refrigerades, i quan hi participen una o quatre substàncies. El tipus d'accident amb més morts és l'explosió i el núvol de gas. En quant al tipus d'activitat és l'originat en plantes de procés o en plantes d'emmagatzematge. Les causes amb més morts són una fallida mecànica i una errada humana.

En canvi, els accidents que tenen un nombre menor de morts són els que s'han produït a la dècada dels 50 o després dels anys 70 i en estats desenvolupats. Els accidents amb pocs morts es tenen també quan les substàncies que intervenen estan en estat físic gas o gas pressuritzat, quan són corrosives, i quan hi participen dues substàncies. El tipus d'accident que provoca menys morts és la fuga i l'incendi. Pel que fa al tipus d'activitat el que causa menys morts és l'originat durant tasques auxiliars de la indústria, i les causes amb menys morts són l'impacte i la fallida reacció violenta.

Pel que fa al nombre de ferits, els accidents que tenen un major nombre de ferits són els produïts a la dècada dels anys 40, abans de la dècada dels 70, i en els països en via de desenvolupament. En els accidents amb molts ferits les substàncies que intervenen es troben en estat físic sòlid o líquid, són tòxiques i explosives, i quan hi participen quatre substàncies. El tipus d'accident amb més ferits és el núvol de gas i la fuga. Pel que fa al tipus d'activitat el que provoca més ferits és l'originat durant operacions de càrrega i descàrrega o durant tasques auxiliars de la indústria. Les causes amb més ferits són per una reacció violenta i una errada humana.



En canvi, els accidents que tenen un nombre menor de ferits són els que s'han produït a la dècada dels 70, en general després dels anys 70, i en estats desenvolupats. Quan les substàncies que intervenen estan en estat físic gas, són asfixiants, i quan hi participen dues substàncies. El tipus d'accident amb menys ferits és l'incendi. Pel que fa al tipus d'activitat és l'originat en locals domèstics o comercials o en plantes d'emmagatzematge. I per últim la causa que provoca menys ferits és una variació en les condicions de procés.

I per finalitzar, l'estimació econòmica del dany material actualitzada amb l'índex Marshall & Stevens, és més elevada quan els accidents s'han produït a la dècada dels anys 80, després de la dècada dels 70, i en els països subdesenvolupats. Els danys també són elevats quan les substàncies que intervenen estan en estat físic gas o pols, són refrigerades, i quan hi participen quatre substàncies o una. El tipus d'accident amb més danys és la fuga, i pel que fa al tipus d'activitat la més costosa és l'originada en plantes de procés o en plantes d'emmagatzematge. Les causes que produeixen més danys són per una errada humana i per una fallida d'instruments.

Per contra, els accidents que tenen una estimació econòmica menor són els que s'han produït a la dècada dels 50, en general abans dels anys 70, i en estats en via de desenvolupament. Quan les substàncies que intervenen estan en estat físic sòlid o líquid, són oxidants, i quan hi participen dues substàncies. El tipus d'accident amb menys danys és el núvol de gas i l'incendi. Pel que fa al tipus d'activitat és l'originat durant tasques auxiliars de la indústria. I per últim la causa que provoca un cost menor és una fallida de serveis.

# CAPÍTOL 5: MODELS LINEALS

## 5.1. Introducció

Una variable resposta  $y$ , depèn de moltes altres variables  $x_1, \dots, x_n$  encara que algunes d'aquestes poden ser no observables o, fins i tot, desconegudes per l'investigador. El model de regressió pretén capturar l'efecte de les variables més importants, i representa l'efecte de les restants mitjançant una variable aleatòria que anomenarem pertorbació del model.

El model general de regressió es basa en les següents hipòtesis bàsiques:

1. Pel conjunt fix de valors de les  $x$ , la distribució de  $y$  té per mitjana :

$$E[y | x_1, \dots, x_{p-1}] = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1}$$

2. La variància de  $y$  és constant i no depèn dels valors de les  $x$ :

$$\text{Var}(y | x_1, \dots, x_{p-1}) = \sigma^2$$

3. Les variables  $y_i$  són independents entre si;
4. La variable resposta es distribueix normalment.

Aquestes 4 hipòtesis es poden resumir de forma matricial a través de:

$$\underline{Y} \sim N(\underline{X}\beta, \sigma^2 \underline{I}) \text{ on}$$

$$\underline{\underline{X}} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p-11} \\ 1 & x_{12} & x_{22} & \dots & x_{p-12} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 1 & x_{1n} & x_{2n} & \dots & x_{p-1n} \end{bmatrix} \quad \underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_{p-1} \end{bmatrix}$$

i  $\underline{I}$  és la matriu  $n \times n$  identitat.

A més a més, d'aquestes condicions, se suposa que es disposa de més observacions en punts  $\underline{X}$  diferents que paràmetres estimats. Com que el model inclou  $p$  paràmetres  $\beta$  això obliga a que:

5. Cap de les variables explicatives sigui una combinació lineal de la resta. Les columnes de la matriu  $\underline{X}$  han de ser linealment independents i per tant,  $\text{Rang}(\underline{X}) = p$  de manera que el nombre de files sigui major que  $p$ .

Per més informació es pot consultar: Daper i Smith (1.998), Peña (1987), o Weisberg (1.985)

Cal remarcar que el model lineal de regressió suposa normalitat per a la variable resposta. En aquest estudi, clarament dues de les tres variables resposta no segueixen una distribució normal, ja que són discretes. Aquest és el cas de les variables **KILLED** i **INJURED**. Aquestes variables són comptetjos i s'haurien de modelar mitjançant models lineals generalitzats, on la variable resposta segueix una distribució de Poisson. Una bona referència sobre aquest models és Mc Cullagh i Nelder (1.987). En compte de fer-ho així, es farà aproximant aquests models mitjançant models lineals.

Sovint els models lineals es poden adaptar per aquest tipus de dades a base de transformar la variable resposta. En particular, quan  $\text{Var}(y) = k E[y]$ , com és el cas de la Poisson, ( $\text{Var}(y) = E[Y]$ ), prendre logaritmes estabilitza la variància i sovint linealitza el model. Per tant, s'ha modelat el logaritme neperià de les variables

**KILLED** i **INJURED** fent servir models lineals. Aplicar logaritmes és el millor que es pot fer, en aquest cas, quan en comptes d'utilitzar un model lineal generalitzat s'utilitza un model lineal de regressió.

En canvi, la tercera variable resposta, **DAMAGE**, no presenta aquest problema, ja que el logaritme neperià d'aquesta variable té un comportament molt continu (veure capítol 4).

Les proves d'hipòtesis per eliminar del model aquelles variables no significatives es fan formulant una hipòtesi nul·la i una alternativa. Es calcula l'estadístic  $F$  definit a continuació, de manera que quan la  $H_0$  és certa, aquest es distribueix d'acord amb una  $F(q, N-p)$  de Fisher. Per exemple si es vol fer la següent prova d'hipòtesi per la variable **ESTAT\_FÍSIC**:

$$H_0 : \beta_{Efsolid} = \beta_{Efliquid} = \beta_{Efgas} = \beta_{Efpigas} = \beta_{Efdust} = 0$$

$$H_1 : \neg H_0$$

$$F = \frac{(SQ_{RO} - SQ_R)}{\frac{q}{S_R^2}} \sim F(q, N-p)$$

on  $SQ_{RO}$  és la suma de quadrats dels residus pel model quan la  $H_0$  és certa,  $SQ_R$  és la suma de quadrats dels residus quan la  $H_1$  és certa,  $S_R^2$  és la variància residual del model sota  $H_1$ ,  $q$  és el nombre de restriccions de  $H_0$  i  $(N-p)$  són els graus de llibertat de l'error residual sota  $H_1$ .

Si  $F < F^{\alpha/2}(q, N-p)$ , s'accepta  $H_0$  i s'eliminen del model les variables no significatives d'acord amb  $H_0$ . En aquest exemple concret, si s'accepta  $H_0$  es treuen del model les variables indicadores de la variable **ESTAT\_FÍSIC**.

## 5.2. Taula de les variables resposta

D'acord amb el que s'ha descrit als capítols anteriors, les variables resposta són les que es mostren a la Taula 31. Cal destacar la importància que els missings tenen en aquestes variables, ja que fan reduir, en un nombre important, el nombre d'observacions on es coneix el valor de la resposta.

El rang de variació és molt gran i això farà difícil trobar un model que ajusti bé tots els accidents. Hi haurà residus estandarditzats grans que tindran molta influència sobre la recta ajustada, però “se la mereixen”, ja que no són errors.

A l'hora de construir els models, per les raons comentades anteriorment, s'utilitzarà el logaritme neperià de les dades. Tot i fer la transformació logarítmica per a la variable **EVACUATED**, no s'ha aconseguit un resultat satisfactori, perquè els accidents amb pocs o cap evacuat continuen tenint molta influència sobre la resta i no s'ha aconseguit normalitat. Per tant, d'acord amb l'expert responsable del projecte s'ha decidit prescindir d'aquesta variable com a resposta.

| VARIABLE  | RANG                          | ZEROS | MISSINGS | NO MISSINGS | TOTAL |
|-----------|-------------------------------|-------|----------|-------------|-------|
| KILLED    | 0-2.000<br>persones           | 1.037 | 3.024    | 2.143       | 5.167 |
| INJURED   | 0-9.999<br>persones           | 779   | 2.374    | 2.793       | 5.167 |
| EVACUATED | 0-200.000<br>persones         | 276   | 4.055    | 1.112       | 5.167 |
| DAMAGE    | 0,002-456,53<br>milions de \$ | 0     | 4.417    | 750         | 5.167 |

Taula 31: Resum de les variables resposta.

### **5.3. Taules de les variables explicatives**

A continuació es mostra la Taula 32 amb les variables categòriques escollides per utilitzar-les com a explicatives. Aquestes variables són : **ESTAT FÍSIC**, **RISC DE LA SUBSTÀNCIA**, **TIPUS D'ACCIDENT**, **TIPUS D'ACTIVITAT**, **CAUSA GENERAL** i **NIVELL DE DESENVOLUPAMENT**.

Inicialment, la base de dades estava formada per moltes més variables que es van descartar en el seu dia. Per exemple, hi havien variables que presentaven moltes dades mancants i que per aquest motiu es van descartar, és el cas de **PD**, **IS**, **IG** i **SC**. D'altres van ser descartades per tenir massa categories, com és el cas de les variables **LOCATION**, **REGION**, **COUNTRY**, **MN**, **MC**, **IT2**, **IS**, **OG2** i **SC**. Com què, en general, es disposava de variables que permetien descriure un succés de forma més general, es va decidir no tenir-les en compte. Les variables que indiquen la data de l'accident (**DIA**, **MES** i **ANY**) i les variables de text **AB**, **KW** i **RA**, i la categòrica **CR** també van ser descartades.

La majoria de variables explicatives tenen categories no exloents, és a dir, que un mateix registre (accident) pot tenir més d'una categoria de la variable. Per exemple, una substància pot ser alhora explosiva i tòxica. Les variables amb aquesta característica són: **ESTAT\_FÍSIC**, **HAZARD** (risc de la substància), **TIPUSACC** (tipus d'accident) i **CAUSAG** (causa general). A partir de les diferents categories de les variables explicatives s'han creat variables indicadores, amb un nom identificatiu de la variable d'origen i la categoria.

Un altre tema important a tenir en compte és el dels missings en alguna d'aquestes variables. Aquest és un greu problema a tenir en compte en el modelat de les variables resposta perquè les dades mancants redueixen el nombre d'observacions útils amb les que es construirà el model.

| VARIABLE                             | EXCLOENT | CATEGORIES  | MISSINGS | NO MISSINGS | TOTAL |
|--------------------------------------|----------|---|----------|-------------|-------|
| Estat Físic (Ef*)                    | NO       | Sòlid (Efsòlid)<br>Líquid (Efliquid)<br>Gas (Efgas)<br>Gas pressuritzat (Efplgas)<br>Dust (Efdust)  | 599      | 4.568       | 5.167 |
| Hazard “Risc de la Substància” (HZ*) | NO       | Tòxic (HZTO)<br>Inflamable (HZFI)<br>Explosiu (HZEX)<br>Corrosiu (HZCO)<br>Radioactiu (HZRA)<br>Refrigerat (HZCD)<br>Asfixiant (HZAS)<br>Oxidant (HZOX)   | 0        | 5.167       | 5.167 |
| Tipus Accident (Tacc*)               | NO       | Explosió (Taccex)<br>Incendi (Taccfi)<br>Núvol de gas (Taccgac)<br>Fuga (Taccfe)  | 205      | 4.962       | 5.167 |
| Tipus Activitat (Tact*)              | SI       | Locals domèstics o comercials (Tactdo/c)<br>Planta de procés (Tactpro)<br>Planta d'emmagatzematge (Tactsto)<br>Càrrega/Descàrrega (Tacttrans)<br>ALTRES (Tactalt)                                 | 185      | 4.982       | 5.167 |
| Causa General (Cg*)                  | NO       | Externa (Cgext)<br>Humana (Cghum)<br>Impacte (Cgimp)<br>Mecànica (Cgmec)<br>Reacció violenta (Cgvrea)<br>Instrument (Cginst)<br>Variació en les condicions de procés (Cgproc)<br>Serveis (Cgserv) | 2.562    | 2.605       | 5.167 |
| Nivell de desenvolupament (M*)       | SI       | Estats Desenvolupats (Mdesenv)<br>E. en via de desenvolupament (Menviadesenv)<br>Estats Subdesenvolupats (Msubdesenv)   | 0        | 5.167       | 5.167 |

Taula 32: Resum de les variables explicatives categòriques.

La Taula 33 conté les variables quantitatives que s'utilitzaran com a explicatives. Cal destacar el nombre important de missings que presenta la variables **QSUS** (quantitat de substància), i per tant **LNQSUS**, cosa que dificultarà la seva inclusió als models perquè reduirà dràsticament el nombre de graus de llibertat, per construir el model.

Pel que fa a la variable nombre de substàncies (**NSUS**), es tractarà com si fos una variable continua, ja que la seva categorització segueix un ordre creixent.

| VARIABLE                       | RANG         | ZEROS | MISSINGS | NO MISSINGS | TOTAL |
|--------------------------------|--------------|-------|----------|-------------|-------|
| Nombre de Substàncies (NSUS)   | 1-6          | -     | 0        | 5.167       | 5.167 |
| Quantitat de Substància (QSUS) | 0,001-99.999 | 0     | 4.026    | 1.141       | 5.167 |

Taula 33: Resum de la variables explicatives quantitatives.

En un principi, es partirà sempre d'un model amb les variables indicadores (0/1): de la variable estat físic (**Efsòlid**, **Eflíquid**, **Efgas**, **Efplgas** i **Efdust**), del risc de la substància (**HZTO**, **HZFI**, **HZEX**, **HZCO**, **HZRA**, **HZCD**, **HZAS** i **HZOX**), del tipus d'accident (**Taccex**, **Taccfi**, **Taccgac** i **Taccree**), del tipus d'activitat (**Tactdo/c**, **tactpro**, **tactsto** i **Tacttrans**), de la causa general (**Cgext**, **Cghum**, **Cgum**, **Cgmec**, **Cgvrea**, **Cginst**, **Cgproc** i **Cgserv**) i del nivell de desenvolupament (**Mdesenv** i **Menvia**); les variables quantitatives **NSUS** i **LNQSUS** i els productes entre elles. Però, quan alguna explicativa tingui com a missings més de la meitat d'observacions, es prescindirà d'ella.



## **5.4. Estratègia a seguir en el modelat**

Quan una interacció (producte de variable categòrica per quantitativa) és significativament diferent de zero, s'ha de conservar al model els efectes principals (variable individual que intervé en el producte), i això no ho respecta la regressió pas a pas. Aquesta estratègia que consisteix en afegir i eliminar les variables seqüencialment intentant que a cada pas el model millori va bé quan no hi ha interaccions i les variables explicatives són contínues, ja que la regressió pas a pas no té en compte la jerarquia dels productes de les variables explicatives (les numèriques per les indicadores).

Com que gairebé totes les variables explicatives són categòriques, el que s'ha fet ha estat establir un algorisme, per decidir en quin ordre s'eliminen les variables. L'ordre en què es treuen del model les variables no significatives condiciona el model final al qual s'arriba. Cal recordar que quan es treu una variable no significativa canvien els coeficients, les desviacions tipus i els t-ratio dels coeficients de les variables que es queden al model, i això fa que aquestes variables es tornin més o menys significatives. Això passa sempre que les dades són observacionals, com és el cas de les dades de MHIDAS.

D'acord amb l'enginyer responsable del projecte s'ha construït el següent algorisme tenint en compte l'ordre d'importància de les variables explicatives:

1. Es fa la prova d'hipòtesi per veure si tots els coeficients dels productes de la variable **NSUS** (nombre de substàncies) per la resta de variables indicadores són zero.
2. Es fa el mateix per la variable **LNQSUS** (logaritme neperià de la quantitat de substància).
3. Si els coeficients dels productes de la variable **NSUS** no són significativament diferents de zero, passar al punt 6.

4. Els coeficients dels productes són significativament diferents de zero, no es poden treure tots de cop, i s'intenta treure'ls per subgrups.
  - 4.1. Es fa la prova d'hipòtesi pels productes de la variable **NSUS** per la variable **CAUSA GENERAL**.
  - 4.2. Es fa la prova d'hipòtesi pels productes de la variable **NSUS** per la variable **ESTAT FÍSIC**.
  - 4.3. Es fa la prova d'hipòtesi pels productes de la variable **NSUS** per la variable **TIPUS D'ACCIDENT**.
  - 4.4. Es fa la prova d'hipòtesi pels productes de la variable **NSUS** per la variable **TIPUS D'ACTIVITAT**.
  - 4.5. Es fa la prova d'hipòtesi pels productes de la variable **NSUS** per la variable **HAZARD**.
  - 4.6. Es fa la prova d'hipòtesi pels productes de la variable **NSUS** per la variable **NIVELL DE DESENVOLUPAMENT**.
5. Si els productes de la variable **LNQSUS** són significatius fer l'apartat 4 amb aquesta variable.
6. Fer la prova d'hipòtesi per veure si el coeficient de la variable **NSUS** és zero.
7. Fer la prova d'hipòtesi per veure si el coeficient de la variable **LNQSUS** és zero.
8. Fer la prova d'hipòtesi per les variables **CG\***.
9. Fer la prova d'hipòtesi per les variables **EF\***.
10. Fer la prova d'hipòtesi per les variables **TACC\***.
11. Fer la prova d'hipòtesi per les variables **TACT\***.
12. Fer la prova d'hipòtesi per les variables **HZ\***.
13. Fer la prova d'hipòtesi per les variables **M\***.

En el cas de no acceptar que totes les variables indicadores d'una variable siguin zero, però s'observa que un subconjunt si que ho és, s'ha fet la prova d'hipòtesis per aquest subconjunt. En fer les proves d'hipòtesis tenim un 5% de possibilitats de rebutjar o acceptar la hipòtesis nul·la erròniament. Això pot fer que no s'inclogui una variable important o que en el model final n'hi hagi una que no sigui important per explicar la resposta.

Un cop s'ha arribat al punt 13, el següent pas és treure d'una en una les variables explicatives no significatives fins que no hi hagi cap que ho sigui, començant per la variable que té el p-valor més gran. Es considera que una variable no és significativa quan té un coeficient amb un p-valor més gran que 0,05.

# CAPÍTOL 6: MODEL LINEAL PER A KILLED

## 6.1. Introducció

En aquest capítol es construiran models lineals pel logaritme neperià del nombre de morts en accidents industrials amb substàncies perilloses, és a dir, per a la variable **LNKILLED**. Se separen els accidents que tenen com a mínim una víctima mortal dels que no en tenen cap. Això es fa degut a que gairebé la meitat dels accidents, on es coneix el nombre de morts, no tenen cap víctima mortal. D'aquesta manera es comprovarà si els accidents que causen morts s'expliquen amb variables diferents o amb les mateixes que expliquen tots els accidents (comparant ambdós models).

Quan s'analitzen tots els accidents hi ha un efecte que no es desitja i és que es treballa amb dues poblacions diferents. Els accidents amb morts no són una mostra representativa de la població amb tots els accidents. A la base de dades hi són tots els accidents on hi ha hagut algun mort, però no tots els accidents on no hi ha hagut cap mort. I això pot esbiaixar els resultats.

Però cal recordar que hi ha variables que contenen un gran nombre de missings, tal i com s'observa a les taules del capítol 5. Això representa un greu problema a l'hora de construir un model lineal, ja que fa disminuir el nombre d'observacions útils amb les quals es construirà el model.

La variable **KILLED** presenta 3.024 missings. Per aquest motiu s'ha creat una nova base de dades eliminant els accidents en els que es desconeixen el nombre de morts, amb la qual cosa es disposa d'un total de 2.143 accidents. Un cop feta aquesta selecció, cal veure quines variables explicatives tenen prou observacions i quines no.

## 6.2. Models pels accidents amb morts

Per construir el model pels accidents que tenen com a mínim un mort i sense missings, s'ha creat una nova base de dades eliminant els 1.037 accidents que no tenen cap mort dels 2.143 que no tenen missings a la variable **KILLED**. Amb la qual cosa es disposa de 1.106 accidents. La variable **QSUS**, tal i com s'observa a la taula 34, té 875 missings, i com aquest fet redueix dràsticament el nombre de graus de llibertat no entrarà en el model, ni el logaritme neperià de la quantitat de substància (**LNQSUS**) ni els productes d'aquesta variable per les diferents variables indicadores. Les causes generals (**CG\***) també presenten un nombre important de missings, 554. S'ha decidit entrar-la en el model encara que redueix gairebé a la meitat els graus de llibertat disponibles, però si en el model final hi ha alguna causa general, aleshores es construirà un altre model sense tenir en compte aquesta variable. Aquest últim model tindrà una mica més del doble d'observacions i això pot fer que el model final sigui un altre.

| VARIABLE                       | MISSINGS |
|--------------------------------|----------|
| ESTAT FÍSIC (EF*)              | 90       |
| RISC DE LA SUBSTÀNCIA (HZ*)    | 0        |
| NOMBRE DE SUBSTÀNCIES (NSUS)   | 0        |
| TIPUS D'ACCIDENT (TACC*)       | 36       |
| TIPUS D'ACTIVITAT (TACT*)      | 35       |
| CAUSA GENERAL (CG*)            | 554      |
| NIVELL DE DESENVOLUPAMENT (M*) | 0        |
| QSUS                           | 875      |

Taula 34: Nombre de missings de les variables explicatives.  
El nombre total d'accidents amb morts és 1.106.

### 6.2.1. Model amb la variable CAUSA GENERAL

El model final al qual s'arriba amb les variables estat físic (EF\*), risc de la substància (HZ\*), nombre de substàncies (NSUS), tipus d'accident (TACC\*), tipus d'activitat (TACT\*), causa general (CG\*), nivell de desenvolupament de l'estat (M\*) i els productes del nombre de substàncies amb la resta de variables indicadores (NS•\*) té les següents observacions amb un residu més gran que 3 en valor absolut:

| Obs  | Eflíquid | lnKilled | Fit    | StDev Fit | Residual | St Resid |
|------|----------|----------|--------|-----------|----------|----------|
| 1096 | 0,00     | 4,8760   | 1,1512 | 0,0955    | 3,7247   | 3,40R    |
| 1098 | 1,00     | 5,0311   | 1,6206 | 0,1691    | 3,4105   | 3,14R    |
| 1102 | 0,00     | 6,2148   | 2,3605 | 0,1988    | 3,8543   | 3,57R    |
| 1103 | 0,00     | 6,3299   | 1,6720 | 0,1401    | 4,6579   | 4,27R    |
| 1106 | 1,00     | 7,6010   | 0,9793 | 0,1709    | 6,6216   | 6,10R    |

Quadre 1: Llistat d'anomalies del model pel Logaritme neperià de Killed pels accidents amb morts

Aquestes anomalies es caracteritzen fonamentalment per correspondre totes a accidents amb un nombre elevat de morts. Aquests 5 accidents són:

- L'observació 1.096 té com codi, AN, el 1755, amb 131 morts.
- L'observació 1.098 té com codi, AN, el 989, amb 153 morts (aquest accident ja era outlier en el capítol 4, per tenir un major nombre de morts en relació als accidents de la dècada dels 80 i en relació als accidents produïts per una causa mecànica).
- L'observació 1.102 té com codi, AN, el 420, amb 500 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts en estats en via de desenvolupament (variable que pertany al model final) i com ja s'havia mencionat en el capítol 4 per tenir un gran nombre de morts en relació a la dècada dels 80, als accidents produïts amb substàncies en estat físic gas pressuritzat i inflamables i als accidents produïts en plantes d'emmagatzematge.

- L'observació 1.103 té com codi, AN, el 1668, amb 561 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts en estats desenvolupats, amb substàncies explosives (aquestes variables pertanyen al model final) i com ja s'havia mencionat en el capítol 4 per tenir un gran nombre de morts en relació als accidents en els que intervé 1 substància i que aquesta sigui oxidant, i als accidents produïts en plantes d'emmagatzematge.
- L'observació 1.106 té com codi, AN, el 1098, amb 2.000 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts amb substàncies en estat físic líquid (aquesta variable pertany al model final) i com ja s'havia mencionat en el capítol 4 per tenir un gran nombre de morts en relació als accidents produïts estats subdesenvolupats, en la dècada dels 80, amb 1 substància i que aquesta sigui inflamable i/o tòxica.

Aquestes dues últimes observacions són valors extrems i en ajustar la recta de regressió per mínims quadrats influeixen molt en els pendents i en els coeficients del model ajustat. Però és el que toca, són valors reals allunyats de la resta d'observacions pel seu elevat nombre de morts. Com que totes aquestes anomalies corresponen a accidents grans, es podria pensar en separar els accidents amb molts morts dels que tenen pocs o cap i modelar-los per separat, però no hi ha prou accidents grans per fer-ho amb garanties.

La influència dels valors atípics segurament no se solucionaria a través de l'ajust de models lineals generalitzats, sinó que s'hauria de fer a través de criteris de regressió robusta, que dóna menys importància als valors anòmals. Si no s'ha fet és perquè no es disposa del software apropiat.

Si es torna a ajustar el model sense tenir en compte aquests dos últims outliers el model canvia una mica, si es compara amb el model que els inclou (Quadre 3), no hi ha cap interacció, la **NSUS** no és significativament diferent de zero i per tant no forma part del model, com tampoc **TACCREL**. El model que quedaria és:

$$\ln\text{Killed} = 1,40 - 0,319 \text{ Eflíquid} - 0,338 \text{ HZTO} + 0,638 \text{ HZEX} + \\ + 0,767 \text{ HZCD} + 0,468 \text{ CGEXT} - 0,274 \text{ Mdesenv} + \\ + 0,525 \text{ Menviadesenv}$$

530 cases used 574 cases contain missing values

| Predictor | Coef     | StDev   | T     | P     |
|-----------|----------|---------|-------|-------|
| Constant  | 1,3967   | 0,1393  | 10,03 | 0,000 |
| Eflíquid  | -0,31887 | 0,09523 | -3,35 | 0,001 |
| HZTO      | -0,3384  | 0,1057  | -3,20 | 0,001 |
| HZEX      | 0,6382   | 0,1279  | 4,99  | 0,000 |
| HZCD      | 0,7673   | 0,3994  | 1,92  | 0,055 |
| CGEXT     | 0,4679   | 0,1195  | 3,92  | 0,000 |
| Mdesenv   | -0,2743  | 0,1319  | -2,08 | 0,038 |
| Menviade  | 0,5253   | 0,1890  | 2,78  | 0,006 |

S = 1,034      R-Sq = 20,4%      R-Sq(adj) = 19,4%

Quadre 2: Llistat de Minitab del model final per **LNKILLED** pels accidents amb morts amb la variable **CAUSA GENERAL** i sense 2 outliers.

Tot i que el model millora una mica (si es compara amb el model del Quadre 3), es deixaran aquestes dues observacions perquè no són cap error, aquests accidents han succeït i han tingut un nombre de morts més elevat del que era previsible segons el model i les seves característiques.

$$\ln\text{Killed} = 1,64 - 0,297 \text{ Eflíquid} + 0,041 \text{ HZEX} - 0,102 \text{ NSUS} - \\ - 0,257 \text{ TACCREL} + 0,443 \text{ CGEXT} - 0,382 \text{ Mdesenv} + \\ + 0,384 \text{ Menviadesenv} + 0,480 \text{ ns*hzex}$$

519 cases used 587 cases contain missing values

| Predictor | Coef     | StDev   | T     | P     |
|-----------|----------|---------|-------|-------|
| Constant  | 1,6356   | 0,1860  | 8,80  | 0,000 |
| Eflíquid  | -0,2970  | 0,1044  | -2,84 | 0,005 |
| HZEX      | 0,0408   | 0,2983  | 0,14  | 0,891 |
| NSUS      | -0,10220 | 0,09095 | -1,12 | 0,262 |
| TACCREL   | -0,2570  | 0,1121  | -2,29 | 0,022 |
| CGEXT     | 0,4428   | 0,1295  | 3,42  | 0,001 |
| Mdesenv   | -0,3822  | 0,1412  | -2,71 | 0,007 |
| Menviade  | 0,3843   | 0,2026  | 1,90  | 0,058 |
| ns*hzex   | 0,4800   | 0,2022  | 2,37  | 0,018 |

S = 1,099      R-Sq = 18,5%      R-Sq(adj) = 17,3%

Quadre 3: Llistat de Minitab del model final per **LNKILLED** pels accidents amb morts i amb la variable **CAUSA GENERAL** amb els dos outliers.



El model final per la variable **KILLED** és l'exponencial del model del Quadre 3 menys 0,1 (ja que per evitar la indeterminació del  $\ln(0)$ , es va sumar 0,1 a la variable **KILLED**).

Tot i que la variable **NSUS** i **HZEX** no són significativament diferent de zero, s'han deixat en el model final perquè la interacció **NS\*HZEX** és significativa.

Abans de decidir-se per la logarítmica s'han provat altres transformacions, com per exemple l'arrel quadrada de la variable **KILLED**, però els millors resultats s'obtenen amb la transformació escollida.

A la Figura 59 s'observa el gràfic de normalitat dels residus del model final pel logaritme neperià de la variable **KILLED**, com ja es va comentar en el capítol 5, aquests residus no segueixen una distribució del tot normal.

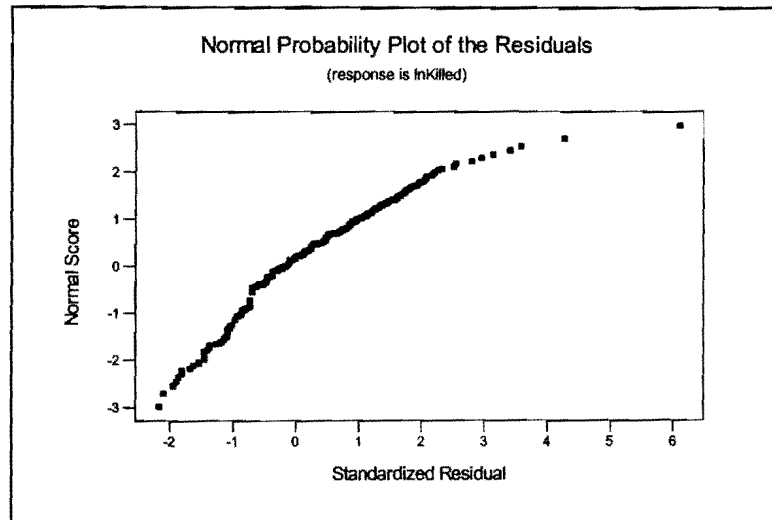


Figura 59: Gràfic de normalitat dels residus del model final pel **LNKILLED** pels accidents amb morts i amb la variable **CAUSA GENERAL**.

A la Figura 60 es poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de **KILLED** i els mateixos gràfics pel nombre de substàncies. En aquests s'hi detecten algunes anomalies, ja detectades abans, i queda clar el fet que els residus no acaben de ser aleatoris, ja que la variable **KILLED** és discreta. En els gràfics dels valors previstos es diferencien els accident amb un mort de la resta, ja que hi ha molts accidents que tenen un mort.

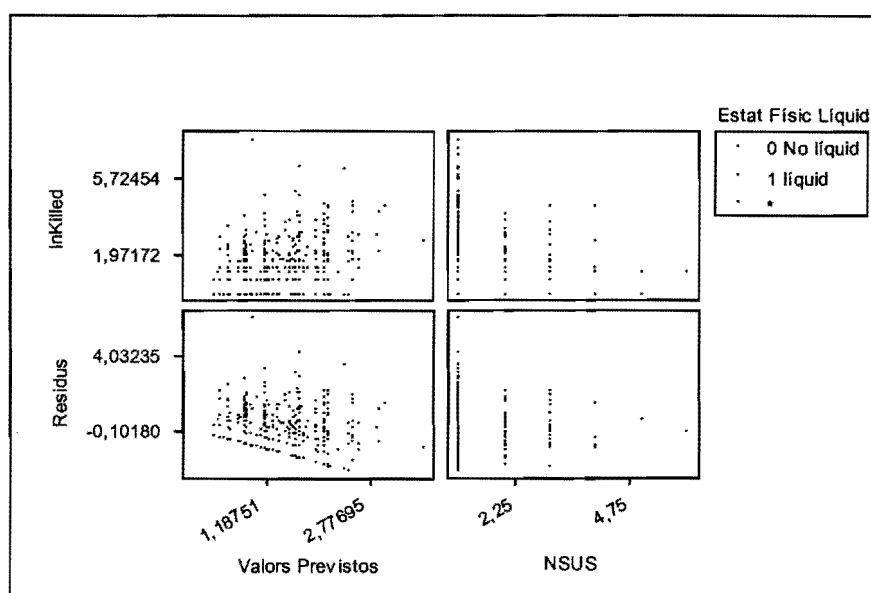


Figura 60: Draftman Plot del model final del **LNKILLED** pels accidents amb morts i amb la variable **CAUSA GENERAL**.

### 6.2.2. Model sense la variable **CAUSA GENERAL**.

El model final al qual s'arriba amb les variables estat físic (**EF\***), risc de la substància (**HZ\***), nombre de substàncies (**NSUS**), tipus d'accident (**TACC\***), tipus d'activitat (**TACT\***), nivell de desenvolupament de l'estat (**M\***) i els productes del nombre de substàncies amb la resta de variables indicadores (**NS.\***) té les següents observacions amb un residu més gran que 3 en valor absolut:

| Obs  | Eflíquid | lnKilled | Fit    | StDev Fit | Residual | St Resid |
|------|----------|----------|--------|-----------|----------|----------|
| 1088 | 1,00     | 4,4079   | 1,0678 | 0,1349    | 3,3401   | 3,14R    |
| 1096 | 0,00     | 4,8760   | 1,7967 | 0,3239    | 3,0793   | 3,01RX   |
| 1097 | 0,00     | 4,8911   | 1,5664 | 0,0835    | 3,3247   | 3,11R    |
| 1098 | 1,00     | 5,0311   | 1,3933 | 0,0769    | 3,6378   | 3,40R    |
| 1099 | 0,00     | 5,7041   | 2,1836 | 0,0789    | 3,5205   | 3,29R    |
| 1100 | 0,00     | 5,7718   | 1,5664 | 0,0835    | 4,2053   | 3,93R    |
| 1101 | 0,00     | 5,8380   | 1,6410 | 0,0779    | 4,1970   | 3,92R    |
| 1102 | 0,00     | 6,2148   | 1,6410 | 0,0779    | 4,5738   | 4,28R    |
| 1103 | 0,00     | 6,3299   | 1,5664 | 0,0835    | 4,7635   | 4,46R    |
| 1104 | 0,00     | 6,9079   | 1,9467 | 0,1593    | 4,9611   | 4,68RX   |
| 1105 | 0,00     | 7,2277   | 2,1836 | 0,0789    | 5,0441   | 4,72R    |
| 1106 | 1,00     | 7,6010   | 1,1424 | 0,1044    | 6,4586   | 6,05R    |

Quadre 4: Llistat d'anomalies del model **LNKILLED** pels accidents amb morts i sense la variable **CAUSA GENERAL**.

Aquestes anomalies es caracteritzen fonamentalment una altra vegada per tenir un nombre elevat de morts. Aquestes són:

- Les observacions 1.096, 1.098, 1.102, 1.103 i 1.106 s'han comentat en l'apartat anterior.
- L'observació 1.088 té com codi, AN, el 1931, amb 82 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts amb substàncies tòxiques (variable que pertany al model final).
- L'observació 1.097 té com codi, AN, el 2401, amb 133 morts. Aquest accident ja havia estat detectat en el capítol 4 per tenir un nombre de morts elevat en relació als accidents produïts per una fallada mecànica.
- L'observació 1.099 té com codi, AN, el 4368, amb 300 morts.
- L'observació 1.100 té com codi, AN, el 2416, amb 321 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts en estats desenvolupats (variable que pertany al model final).
- L'observació 1.101 té com codi, AN, el 8418, amb 343 morts.
- L'observació 1.104 té com codi, AN, el 1886, amb 1.000 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts en estats desenvolupats, amb substàncies explosives (variables que pertanyen al model final).

- L'observació 1.105 té com codi, AN, el 2091, amb 1.377 morts, aquest accident destaca per tenir un major nombre de morts en relació als accidents produïts amb substàncies explosives (variable que pertany al model final).

Gairebé tots aquests accidents ja s'havien destacat al capítol 4 per tenir un nombre elevat de morts en relació a alguna característica diferent a més a més de les mencionades.

El model final pels accidents amb morts i sense la causa general es pot observar al Quadre 5.

| $\ln\text{Killed} = 1,64 - 0,248 \text{ Eflíquid} + 0,380 \text{ Efdust} - 0,251 \text{ HZTO} + 0,543 \text{ HZEX} + 0,773 \text{ HZCD} - 0,617 \text{ Mdesenv}$ |          |         |       |       |
|--|----------|---------|-------|-------|
| 1016 cases used 90 cases contain missing values  |          |         |       |       |
| Predictor  | Coef     | StDev   | T     | P     |
| Constant   | 1,64100  | 0,07786 | 21,08 | 0,000 |
| Eflíquid   | -0,24773 | 0,07412 | -3,34 | 0,001 |
| Efdust   | 0,3803   | 0,1671  | 2,28  | 0,023 |
| HZTO   | -0,25090 | 0,08600 | -2,92 | 0,004 |
| HZEX   | 0,54259  | 0,08908 | 6,09  | 0,000 |
| HZCD   | 0,7729   | 0,3290  | 2,35  | 0,019 |
| Mdesenv  | -0,61717 | 0,07370 | -8,37 | 0,000 |
| S = 1,072      R-Sq = 18,2%      R-Sq(adj) = 17,7%   |          |         |       |       |

Quadre 5: Llistat de Minitab del model final per **LNKILLED** pels accidents amb morts i sense la variable **CAUSA GENERAL**.

Els residus segueixen una distribució semblant a la Figura 59, que correspon al gràfic de normalitat dels residus del model final per tots els morts amb la causa general. I a la Figura 61 es poden observar diferents gràfics, els valors previstos vs residus i el logaritme neperià de Killed, i els mateixos gràfics pel nombre de substàncies. En aquests, es tornen a detectar algunes anomalies, s'observa com els residus no són del tot aleatoris i també com es diferencien els accident amb un mort de la resta, ja que hi ha molts accidents que tenen un mort.

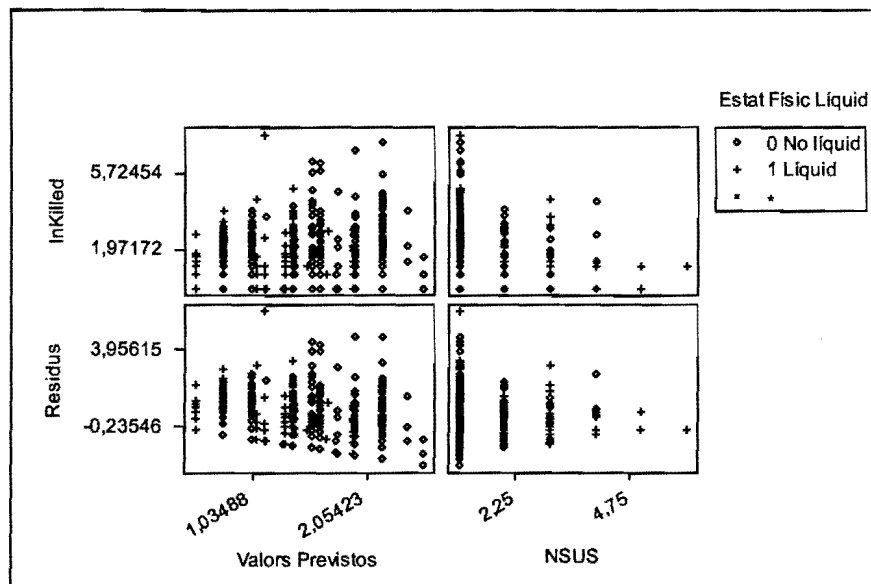


Figura 61: Draftman Plot del model **LNKILLED** pels accidents amb morts i sense la variable **CAUSA GENERAL**.

### **6.3. Interpretació dels models pels accidents amb morts**

A continuació s'interpreten els models trobats pels accidents amb morts, tenint en compte la variable **CAUSA GENERAL** i sense tenir-la en compte. També s'analitzen les dependències entre les variables explicatives que intervenen en el model.

#### **6.3.1. Model amb la variable CAUSA GENERAL**

El model final pels accidents amb morts i incloent la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Killed} = & 1,64 - 0,297 \text{ Eflíquid} + 0,041 \text{ HZEX} - 0,102 \text{ NSUS} - \\ & - 0,257 \text{ TACCREL} + 0,443 \text{ CGEXT} - 0,382 \text{ Mdesenv} + \\ & + 0,384 \text{ Menviadesenv} + 0,480 \text{ ns*hzex} \end{aligned}$$

Quan en l'accident intervenen substàncies en estat físic líquid, el nombre de morts disminueix. El mateix passa quan l'accident es produeix per una fuga i en estats desenvolupats. En canvi, quan la causa que origina l'accident és esdeveniments externs, el nombre de morts augmenta. Això també succeeix quan els accidents es produeixen en estats en via de desenvolupament. Pel que fa a la interacció  $\text{ns*hzex}$ , el que significa és que quan la substància o substàncies que intervenen en l'accident són explosives el logaritme neperià de Killed augmenta en 0,419 per cada substància participant en l'accident.

En analitzar les variables explicatives que formen part del model final es conclou que la variable **MDESENV** té una correlació significativament diferent de zero amb les altres variables explicatives. Per tant aquesta variable s'ha d'interpretar de forma simultània amb la resta d'explicatives. Cal destacar que el 65% dels accidents amb víctimes mortals es produeixen en estats desenvolupats.

Les variables regressores expliquen un 17,3% el nombre de morts en accidents industrials amb substàncies perilloses, això pot ser degut a que hi ha altres variables que expliquen el nombre de morts, però que no hi són a la base de dades MHIDAS.

### 6.3.2. Model sense la variable CAUSA GENERAL

El model final pels accidents amb morts i sense la variable **CAUSA GENERAL** és:

$$\ln\text{Killed} = 1,64 - 0,248 \text{ Eflíquid} + 0,380 \text{ Efdust} - 0,251 \text{ HZTO} + \\ + 0,543 \text{ HZEX} + 0,773 \text{ HZCD} - 0,617 \text{ Mdesenv}$$

Quan en l'accident intervenen substàncies en estat físic líquid el nombre de morts disminueix, però en canvi, quan intervenen substàncies en estat físic pols augmenta. Les substàncies tòxiques contribueixen disminuint el nombre de morts. Però les explosives i les refrigerades fan augmentar-lo. I per últim, els accidents que es produeixen en estats desenvolupats tenen un nombre de morts inferior.

En analitzar les dependències entre variables explicatives s'ha arribat a la conclusió que la variable **EFDUST** està correlacionada amb **HZEX**, ja que més del 68% de les substàncies en estat físic pols són explosives. Per tant, aquestes dues variables s'han d'interpretar conjuntament. És a dir, quan intervenen substàncies explosives i en estat físic pols el nombre de morts augmenta en major grau.

Un altre fet que destaca és el signe negatiu del coeficient de la variable **HZTO**, que indica que quan les substàncies que participen en l'accident són tòxiques el nombre de morts és menor. Això no és lògic des del punt de vista químic, segons l'expert responsable del projecte, però no s'ha trobat cap colinealitat amb les variables que intervenen en el model. Pot ser que hi hagi alguna variable oculta que estigui relacionada amb la variable **HZTO**.

A més a més, com s'ha dit anteriorment, la variable **MDESENV** està correlacionada amb la resta de variables explicatives.

En aquest model, les variables regressores expliquen un 17,7% el nombre de morts en accidents industrials amb substàncies perilloses.



## **6.4. Models per tots els accidents (inclosos els sense morts)**

Per construir el model per a tots els accidents tant els que tenen morts com els que no en tenen, s'ha utilitzat la base de dades amb 2.143 accidents. La variable **QSUS** té 1.539 missings, que són més de la meitat comparant amb el total d'observacions que ara és de 2.143, i com aquest fet redueix dràsticament el nombre de graus de llibertat no es farà servir, ni el logaritme neperià de la quantitat de substància (**LNQSUS**) ni els productes d'aquesta variable per les diferents variables indicadores. Les causes generals (**CG\***) també presenten un nombre important de missings. S'ha decidit entrar-la primer en el model encara que redueix gairebé a la meitat els graus de llibertat (un 40%), i després construir un altre model sense tenir en compte aquesta variable. Aquest últim model tindrà un 40% més d'observacions i això farà que el model final sigui un altre.

| VARIABLE                          | MISSINGS |
|-----------------------------------|----------|
| ESTAT FÍSIC<br>(EF*)              | 131      |
| RISC DE LA SUBSTÀNCIA<br>(HZ*)    | 0        |
| NOMBRE DE SUBSTÀNCIES<br>(NSUS)   | 0        |
| TIPUS D'ACCIDENT<br>(TACC*)       | 56       |
| TIPUS D'ACTIVITAT<br>(TACT*)      | 60       |
| CAUSA GENERAL<br>(CG*)            | 868      |
| NIVELL DE DESENVOLUPAMENT<br>(M*) | 0        |
| QSUS                              | 1.539    |

Taula 35: Nombre de missings de les variables explicatives quan es modelen els morts per tots els accidents. El nombre total d'accidents és 2.143.

### 6.4.1. Model amb la variable CAUSA GENERAL.

El model final al qual s'arriba amb la variable **CAUSA GENERAL** té les següents anomalies amb un residu més gran que 3, en valor absolut:

| Obs  | Eflíquid | lnKilled | Fit     | StDev Fit | Residual | St Resid |
|------|----------|----------|---------|-----------|----------|----------|
| 2114 | 0,00     | 3,9338   | -1,1122 | 0,1337    | 5,0460   | 3,11R    |
| 2133 | 0,00     | 4,8760   | -0,1386 | 0,1412    | 5,0146   | 3,09R    |
| 2140 | 0,00     | 6,3299   | 1,0411  | 0,1879    | 5,2888   | 3,27R    |
| 2143 | 1,00     | 7,6010   | -0,3067 | 0,2138    | 7,9076   | 4,90R    |

Quadre 6: Llistat d'anomalies del model final per **LNKILLED** per tots els accidents i amb la variable **CAUSA GENERAL**.

Aquestes anomalies tornen a ser els accidents amb un nombre elevat de morts. Són:

- Les observacions 2.133 (AN 1755), 2.140 (AN 1668) i 2.143 (AN 1098) s'han comentat en apartats anteriors.
- L'observació 2.144 té com AN el 2317, amb 51 morts, aquest accident destacava al capítol 4 per tenir un major nombre de morts en relació als accidents produïts a la dècada dels 80.

El model final per la variable **KILLED** és l'exponencial del model del Quadre 7 menys 0,1.

| $\ln\text{Killed} = -1,27 - 0,269 \text{ Eflíquid} - 0,287 \text{ HZTO} + 1,18 \text{ HZEX} +$ $+ 0,195 \text{ NSUS} + 0,974 \text{ TACCEX} + 0,975 \text{ TACTDO/C} +$ $+ 0,793 \text{ TACTPRO} + 0,508 \text{ TACTSTO} + 0,791 \text{ TACTTRANS} +$ $+ 0,531 \text{ CGHUM} - 1,08 \text{ Mdesenv} + 0,673 \text{ Menviadesenv}$ |         |         |       |       |
|---|---------|---------|-------|-------|
| 1201 cases used 942 cases contain missing values  |         |         |       |       |
| Predictor   | Coef    | StDev   | T     | P     |
| Constant  | -1,2695 | 0,2791  | -4,55 | 0,000 |
| Eflíquid  | -0,2694 | 0,1002  | -2,69 | 0,007 |
| HZTO  | -0,2868 | 0,1087  | -2,64 | 0,008 |
| HZEX  | 1,1797  | 0,1748  | 6,75  | 0,000 |
| NSUS  | 0,19473 | 0,08681 | 2,24  | 0,025 |
| TACCEX  | 0,9735  | 0,1089  | 8,94  | 0,000 |
| TACTDO/C  | 0,9754  | 0,2662  | 3,66  | 0,000 |
| TACTPRO   | 0,7931  | 0,2055  | 3,86  | 0,000 |
| TACTSTO   | 0,5075  | 0,2090  | 2,43  | 0,015 |
| TACTTRAN  | 0,7915  | 0,2229  | 3,55  | 0,000 |
| CGHUM   | 0,53119 | 0,09781 | 5,43  | 0,000 |
| Mdesenv   | -1,0761 | 0,1626  | -6,62 | 0,000 |
| Menviade  | 0,6728  | 0,2653  | 2,54  | 0,011 |
| S = 1,628      R-Sq = 29,1%      R-Sq(adj) = 28,4%  |         |         |       |       |

Quadre 7: Llistat de Minitab del model final per **LNKILLED** per tots els accidents i amb amb la variable **CAUSA GENERAL**.

A la Figura 62 s'observa el gràfic de normalitat dels residus del model final pel logaritme neperià de la variable **KILLED**. I a la Figura 63 s'hi poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de Killed i els mateixos gràfics pel nombre de substàncies. En ambdós gràfics es detecten anomalies. També s'observa com els residus no acaben de ser aleatoris, ja que la variable **KILLED** és discreta. I els accident que no tenen cap mort es diferencien de la resta molt clarament i els que tenen un mort també, però no tant, ja que hi ha molts accidents que tenen un mort com a màxim.

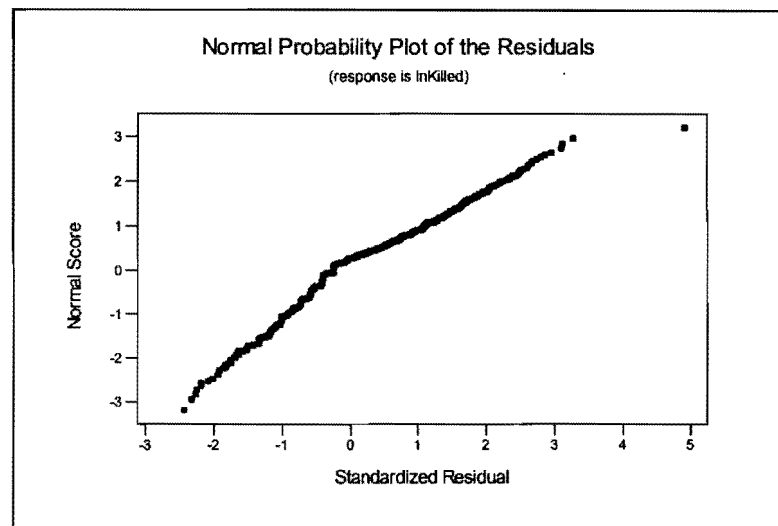


Figura 62: Gràfic de normalitat dels residus del model final del **LNKILLED** per tots els accidents i amb la variable **CAUSA GENERAL**.

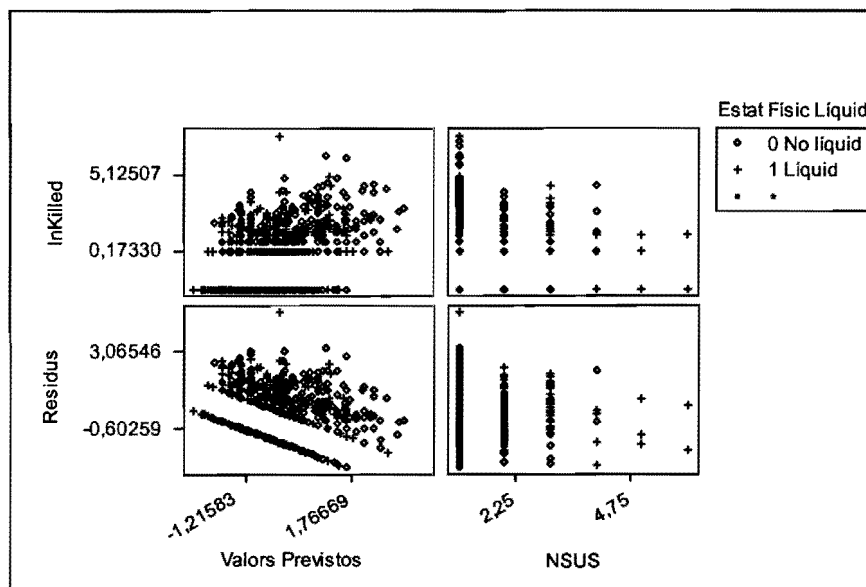


Figura 63: Draftman Plot del model final pel **LNKILLED** pels accidents amb morts i amb la variable **CAUSA GENERAL**.

### 6.4.2. Model sense la variable CAUSA GENERAL.

El model final al qual s'arriba sense la variable **CAUSA GENERAL** té els següents valors anòmals:

| Obs  | Eflíquid | lnKilled | Fit     | StDev Fit | Residual | St Resid |
|------|----------|----------|---------|-----------|----------|----------|
| 2048 | 0,00     | 2,9497   | -1,8636 | 0,1154    | 4,8133   | 3,02R    |
| 2068 | 0,00     | 3,1398   | -1,8636 | 0,1154    | 5,0034   | 3,14R    |
| 2084 | 1,00     | 3,3358   | -1,4793 | 0,0891    | 4,8151   | 3,02R    |
| 2103 | 0,00     | 3,6914   | -1,3268 | 0,1039    | 5,0182   | 3,15R    |
| 2114 | 0,00     | 3,9338   | -1,3268 | 0,1039    | 5,2606   | 3,30R    |
| 2133 | 0,00     | 4,8760   | -0,7522 | 0,1490    | 5,6282   | 3,54R    |
| 2139 | 0,00     | 6,2148   | 1,1776  | 0,1604    | 5,0372   | 3,17R    |
| 2140 | 0,00     | 6,3299   | 0,6949  | 0,1259    | 5,6350   | 3,54R    |
| 2141 | 0,00     | 6,9079   | 0,6949  | 0,1259    | 6,2130   | 3,90R    |
| 2142 | 0,00     | 7,2277   | 2,1549  | 0,1536    | 5,0728   | 3,19R    |
| 2143 | 1,00     | 7,6010   | -0,8953 | 0,1663    | 8,4962   | 5,35R    |

Quadre 8: Llistat d'anomalies del model final pel **LNKILLED** per tots els accidents i sense la variable **CAUSA GENERAL**.

Aquestes anomalies es caracteritzen per tenir un nombre elevat de morts i són:

- Les observacions 2.114 (AN 2317), 2.133 (AN 1755), 2.139 (AN 420), 2.140 (AN 1668), 2.141 (AN 1886), 2.142 (AN 2091) i 2.143 (AN 1098) s'han comentat en apartats anteriors.
- L'observació 2.048 té com AN el 1725, amb 19 morts.
- L'observació 2.068 té com AN el 1905, amb 23 morts.
- L'observació 2.084 té com AN el 814, amb 28 morts.
- L'observació 2.103 té com AN el 2742, amb 40 morts.

El model final per la variable **KILLED** és l'exponencial del model que es mostra al Quadre 9 menys 0,1.

| $\ln\text{Killed} = -0,880 - 0,353 \text{ Eflíquid} - 0,184 \text{ HZTO} + 1,09 \text{ HZEX} +$ $+ 1,63 \text{ HZAS} + 0,185 \text{ NSUS} + 0,927 \text{ TACCEX} - 0,352 \text{ TACCGAC} +$ $+ 0,971 \text{ TACTDO/C} + 0,691 \text{ TACTPRO} + 0,490 \text{ TACTSTO} +$ $+ 0,829 \text{ TACTTRANS} - 1,12 \text{ Mdesenv} + 0,457 \text{ Menviadesenv}$ |          |         |        |       |
|--|----------|---------|--------|-------|
| 1922 cases used 221 cases contain missing values   |          |         |        |       |
| Predictor  | Coef     | StDev   | T      | P     |
| Constant   | -0,8802  | 0,2030  | -4,34  | 0,000 |
| Eflíquid   | -0,35348 | 0,08139 | -4,34  | 0,000 |
| HZTO   | -0,18449 | 0,09503 | -1,94  | 0,052 |
| HZEX   | 1,0947   | 0,1196  | 9,15   | 0,000 |
| HZAS   | 1,6252   | 0,5364  | 3,03   | 0,002 |
| NSUS   | 0,18463  | 0,07243 | 2,55   | 0,011 |
| TACCEX   | 0,92693  | 0,08512 | 10,89  | 0,000 |
| TACCGAC  | -0,3523  | 0,1198  | -2,94  | 0,003 |
| TACTDO/C   | 0,9713   | 0,1918  | 5,06   | 0,000 |
| TACTPRO  | 0,6906   | 0,1539  | 4,49   | 0,000 |
| TACTSTO  | 0,4897   | 0,1589  | 3,08   | 0,002 |
| TACTTRAN   | 0,8289   | 0,1716  | 4,83   | 0,000 |
| Mdesenv  | -1,1209  | 0,1089  | -10,29 | 0,000 |
| Menviade   | 0,4566   | 0,1663  | 2,75   | 0,006 |
| S = 1,597      R-Sq = 33,9%      R-Sq(adj) = 33,5%   |          |         |        |       |

Quadre 9: Llistat de Minitab del model final pel **LNKILLED** per tots els accidents i sense la variable **CAUSA GENERAL**.

La distribució dels residus es semblant a la vista en l'apartat anterior. I a la Figura 64 s'hi poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de Killed i els mateixos gràfics pel nombre de substàncies. Es detecta alguna anomalia, que els residus no acaben de ser aleatoris, ja que la variable **KILLED** és discreta. I com els accident que no tenen morts es diferencien de la resta molt clarament i els que tenen un mort també, però no tant.

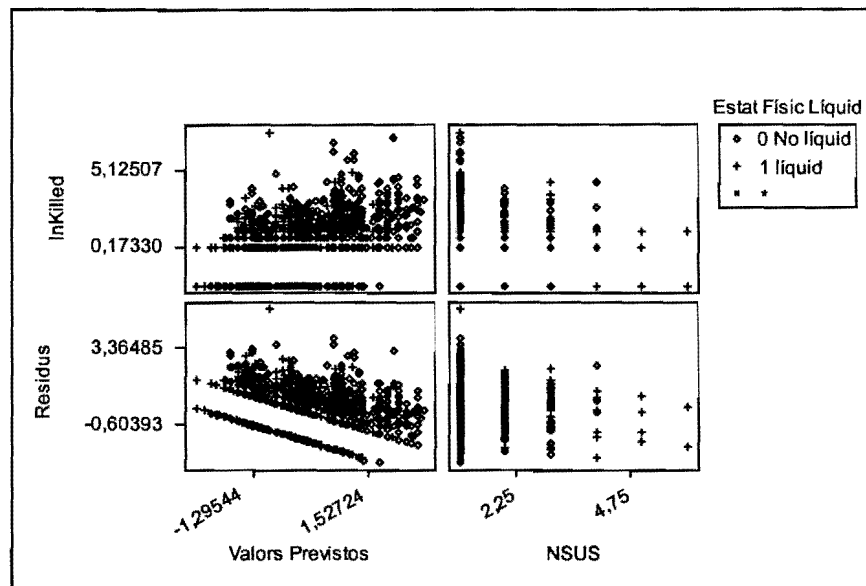


Figura 64: Draftman Plot del model final pel **LNKILLED** de tots els accidents i sense la variable **CAUSA GENERAL**.

## **6.5. Interpretació dels models per tots els accidents (inclosos els sense morts)**

A continuació es mostra la interpretació dels models finals trobats pels accidents amb morts, tenint en compte la variable **CAUSA GENERAL** i sense tenir-la en compte.

### **6.5.1. Model amb la variable CAUSA GENERAL**

El model final per tots els accidents i incloent la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln\text{Killed} = & - 1,27 - 0,269 \text{ Eflíquid} - 0,287 \text{ HZTO} + 1,18 \text{ HZEX} + \\ & + 0,195 \text{ NSUS} + 0,974 \text{ TACCEX} + 0,975 \text{ TACTDO/C} + \\ & + 0,793 \text{ TACTPRO} + 0,508 \text{ TACTSTO} + 0,791 \text{ TACTTRANS} + \\ & + 0,531 \text{ CGHUM} - 1,08 \text{ Mdesenv} + 0,673 \text{ Menviadesenv} \end{aligned}$$

Quan en l'accident intervenen substàncies en estat físic líquid el nombre de morts disminueix. El mateix passa quan intervenen substàncies tòxiques. Quan intervenen substàncies explosives el nombre de morts augmenta. També augmenta a mesura que intervenen més substàncies, quan l'accident es produeix per una explosió, en locals domèstics o comercials, en plantes de procés, plantes d'emmagatzematge, durant operacions de càrrega o descàrrega, quan la causa de l'accident és la humana i quan es produeix en països en via de desenvolupament. I en canvi, disminueix quan l'accident es produeix en estats desenvolupats.

En analitzar les variables explicatives que formen part del model final es conclou que la variable **HZEX** està correlacionada amb **TACCEX**, ja que més d'un 92% de les substàncies explosives participen en accidents produïts per una explosió.



La variable **EFLÍQUID** està correlacionada amb **TACTSTO**, ja que en més d'un 55% dels accidents produïts en plantes d'emmagatzematge intervenen substàncies en estat físic líquid.

La variable **TACTDO/C** està correlacionada amb **TACCEX**, ja que gairebé el 60% dels accidents produïts en locals domèstics o comercials es produeixen per una explosió.

La variable **TACTPRO** està correlacionada amb **TACCEX**, ja que més del 56% dels accidents produïts en plantes de procés s'han originat per una explosió.

La variable **MDESENV** està correlacionada amb la resta de variables explicatives, ja que el 78% dels accidents on es coneix el nombre de morts s'han produït en estats desenvolupats.

També sorprèn que la variable **HZTO** tingui un coeficient negatiu, ja que segons l'expert químic responsable del projecte li correspondria un signe negatiu, però com passava en apartats anteriors, no s'ha trobat cap dependència entre les altres variables explicatives.

Totes les variables que estan correlacionades s'haurien d'interpretar simultàniament.

Les variables regressores expliquen un 28,4% el nombre de morts en accidents industrials amb substàncies perilloses, això és degut a que hi ha altres variables que expliquen el nombre de morts, però que no són a la base de dades MHIDAS.

## 6.5.2. Model sense la variable CAUSA GENERAL

El model final per tots els accidents i sense incloure la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln\text{Killed} = & - 0,880 - 0,353 \text{ Eflíquid} - 0,184 \text{ HZTO} + 1,09 \text{ HZEX} + \\ & + 1,63 \text{ HZAS} + 0,185 \text{ NSUS} + 0,927 \text{ TACCEX} - 0,352 \text{ TACCGAC} + \\ & + 0,971 \text{ TACTDO/C} + 0,691 \text{ TACTPRO} + 0,490 \text{ TACTSTO} + \\ & + 0,829 \text{ TACTTRANS} - 1,12 \text{ Mdesenv} + 0,457 \text{ Menviadesenv} \end{aligned}$$

Quan en l'accident intervenen substàncies en estat físic líquid el nombre de morts disminueix. Això també passa amb les substàncies tòxiques. Però quan intervenen substàncies explosives el nombre de morts augmenta, també passa amb les substàncies asfixiants. El nombre de morts augmenta amb el nombre de substàncies. Quan l'accident es produeix per una explosió, en locals domèstics o comercials, en plantes de procés, en plantes d'emmagatzematge i durant operacions de càrrega i descàrrega el nombre de morts és major. Quan l'accident es produeix per un núvol de gas el nombre de morts disminueix, i quan es produeix en països desenvolupats també. I per últim, el nombre de morts augmenta quan l'accident es produeix en països en via de desenvolupament

En analitzar les variables explicatives que formen part del model final es conclou que a més a més de les variables mencionades abans, estan també correlacionades la variable **TACCGAC** amb **HZTO**, ja que el 80% dels accidents produïts per un núvol de gas participen substàncies tòxiques. I també està correlacionada la variable **HZAS** i **TACCEX**, ja que més del 66% de les substàncies asfixiants han participat en accidents produïts per una explosió. Aquestes variables tenen una correlació significativament diferent de zero i s'han d'interpretar de forma conjunta.

Les variables regressores expliquen un 33,5%. Dels quatre models fets per aquesta variable aquest últim és el que millor explica el nombre de morts en accidents industrials amb substàncies perilloses.

## **6.6. Conclusions**

En analitzar els resultats dels diferents models, es pot concloure que els models per tots els accidents i pels que tenen morts són diferents, tant si es té en compte la variable **CAUSA GENERAL** com si no. Quan es torna a ajustar el model sense tenir en compte la variable **CAUSA GENERAL** hi ha variables que deixen de ser significativament diferent de zero i altres que passen a ser-ho. Cal destacar que en no tenir en compte la variable **CAUSA GENERAL** el model explica més i té menys variabilitat. A més a més, els models amb víctimes mortals tenen menys variabilitat en relació als models per a tots els accidents.

En els quatre models ajustats sempre hi són les variables **Eflíquid**, **HZEX** i **Mdesenv**. La primera variable sempre és negativa, el que significa que quan en l'accident participen substàncies en estat físic líquid el nombre de morts disminueix. El mateix succeeix quan l'accident es produeix en estats en via de desenvolupament. En canvi, quan hi participen substàncies explosives en l'accident el nombre de morts augmenta.

El model trobat per a **KILLED** que més explica és el que ajusta tots els accidents sense tenir en compte la variable **CAUSA GENERAL**, amb una  $R_{adj} = 33,5\%$ . El fet que les variables regressores expliquin poc la variable resposta és degut al soroll de la base de dades i també pot ser degut a que hi ha variables importants que expliquen el nombre de morts i que no estan al nostre abast. Per tant la utilització del model seria més aviat per explicar la variable resposta que per predir-la. També cal destacar que hi ha colinealitat entre algunes variables explicatives.

# CAPÍTOL 7: MODEL LINEAL PER A INJURED

## 7.1. Introducció

En aquest capítol es construïran models lineals pel logaritme neperià del nombre de ferits en accidents industrials amb substàncies perilloses, és a dir, per a la variable **LNINJURED**. Se separen els accidents que tenen com a mínim un ferit dels que no tenen cap. Això es fa degut a que gairebé un 30% dels accidents, on es coneix el nombre de ferits, no en tenen cap.

Amb el nombre de ferits passa el mateix que amb el nombre de morts (capítol 6), es treballa amb dues poblacions diferents. Els accidents amb ferits no són una mostra representativa de la població amb tots els accidents. A la base de dades hi són tots els accidents on hi ha hagut algun ferit, però no tots els accidents on no hi ha hagut cap ferit. I això pot esbiaixar els resultats.

També cal recordar que hi ha variables que contenen un gran nombre de missings, tal i com s'observa a les taules del capítol 5. Això representa un greu problema a l'hora de construir un model lineal, ja que fa disminuir el nombre d'observacions útils amb les quals es construirà el model.

La variable **INJURED** presenta 2.374 missings. Per aquest motiu s'ha creat una nova base de dades eliminant els accidents en els que es desconeixen el nombre de ferits, amb la qual cosa es disposa d'un total de 2.793 accidents.

## **7.2. Models pels accidents amb ferits**

Per construir el model pels accidents que tenen com a mínim un ferit i sense missings, s'ha creat una nova base de dades eliminant els 779 accidents que no tenen cap ferit dels 2.793 que no tenen missings a la variable **INJURED**. Amb la qual cosa es disposa de 2.014 accidents. La variable **QSUS**, tal i com s'observa a la taula 36, té 1.624 missings, i com aquest fet redueix dràsticament el nombre de graus de llibertat no entrarà en el model, ni el logaritme neperià de la quantitat de substància (**LNQSUS**) ni els productes d'aquesta variable per les diferents variables indicadores. Les causes generals (**CG\***) també presenten un nombre important de missings, 995. S'ha decidit entrar-la en el model encara que redueixi gairebé a la meitat els graus de llibertat disponibles però, si en el model final hi ha alguna causa general aleshores es construirà un altre model sense tenir en compte aquesta variable. Aquest últim model tindrà una mica més del doble d'observacions i això pot fer que el model final sigui un altre.

| VARIABLE                          | MISSINGS |
|-----------------------------------|----------|
| ESTAT FÍSIC (EF*)                 | 206      |
| RISC DE LA SUBSTÀNCIA (HZ*)       | 0        |
| NOMBRE DE SUBSTÀNCIES<br>(NSUS)   | 0        |
| TIPUS D'ACCIDENT (TACC*)          | 59       |
| TIPUS D'ACTIVITAT (TACT*)         | 57       |
| CAUSA GENERAL (CG*)               | 995      |
| NIVELL DE DESENVOLUPAMENT<br>(M*) | 0        |
| QSUS                              | 1.624    |

Taula 36: Nombre de missings de les variables explicatives quan es modela els accidents amb ferits. El nombre total d'accidents amb ferits és 2.014.

### 7.2.1. Model amb la variable CAUSA GENERAL.

El model final al qual s'arriba amb les variables estat físic (EF\*), risc de la substància (HZ\*), nombre de substàncies (NSUS), tipus d'accident (TACC\*), tipus d'activitat (TACT\*), causa general (CG\*), nivell de desenvolupament de l'estat (M\*) i els productes del nombre de substàncies amb la resta de variables indicadores (NS\*) té les següents observacions amb un residu més gran que 3 en valor absolut:

| obs  | Efsolid | lnInjure | Fit    | StDev Fit | Residual | St Resid |
|------|---------|----------|--------|-----------|----------|----------|
| 311  | 1,00    | 0,0953   | 4,6829 | 0,4770    | -4,5876  | -3,29RX  |
| 2009 | 0,00    | 7,6010   | 2,8031 | 0,1102    | 4,7978   | 3,27R    |
| 2010 | 0,00    | 7,8241   | 3,0377 | 0,2308    | 4,7864   | 3,29R    |
| 2012 | 0,00    | 8,6995   | 2,2040 | 0,1115    | 6,4955   | 4,42R    |
| 2013 | 0,00    | 9,2103   | 2,2300 | 0,1311    | 6,9803   | 4,76R    |
| 2014 | 0,00    | 9,2103   | 3,2613 | 0,2286    | 5,9490   | 4,09R    |

Quadre 10: Llistat d'anomalies del model **LNIJURED** pels accidents amb ferits i amb la variable **CAUSA GENERAL**.

Aquestes anomalies es caracteritzen fonamentalment per correspondre totes, a excepció de la primera, a accidents amb un nombre elevat de ferits. Aquests 6 accidents són:

- L'observació 311 té com codi, AN, el 8393, amb 1 ferit. Aquest accident és atípic perquè només ha participat una substància sòlida i es va produir en un estat en via de desenvolupament i per tant li correspondria un nombre de ferits més gran segons el model trobat (108,90) i només en té 1.
- L'observació 2.009 té com codi, AN, el 6324, amb 2.000 ferits (aquest accident ja era outlier en el capítol 4 per tenir un major nombre de ferits en relació als accidents de la dècada dels 90).
- L'observació 2.010 té com codi, AN, el 420, amb 2.500 ferits, aquest accident destaca per tenir un major nombre de ferits en relació als accidents produïts en estats en via de desenvolupament (variable que pertany al model final) i per altres característiques mencionades al capítol 4.

- L'observació 2.012 té com codi, AN, el 810, amb 6.000 ferits, aquest accident destaca per tenir un major nombre de ferits en relació als accidents produïts per una errada humana i amb substàncies en estat físic líquid.
- L'observació 2.013 té com codi, AN, el 1098, amb 9.999 ferits, aquest accident destaca per tenir un major nombre de ferits en relació als accidents produïts amb substàncies inflamables (aquesta variable pertany al model final) i com ja s'havia mencionat en el capítol 4 per tenir un gran nombre de ferits en relació als accidents produïts en estats subdesenvolupats, amb substàncies en estat físic líquid i per haver-se produït per una errada humana i una reacció violenta.
- L'observació 2.014 té com codi, AN, el 2744, amb 9.999 ferits, aquest accident destaca per tenir un major nombre de ferits en relació als accidents produïts en estats en via de desenvolupament (aquesta variable pertany al model final) i com ja s'havia mencionat en el capítol 4 per tenir un gran nombre de ferits en relació als accidents produïts amb substàncies en estat físic líquid i per haver-se produït per una errada humana.

```
lnInjured = 2,22 + 0,436 Efsolid - 0,573 HZFI - 0,705 HZEX -
            - 0,0112 NSUS + 0,599 TACCGAC + 0,350 CGEXT +
            + 1,06 Menviadesenv + 0,762 ns*hzex
```

935 cases used 1079 cases contain missing values

| Predictor | Coef     | StDev   | T     | P     |
|-----------|----------|---------|-------|-------|
| Constant  | 2,2152   | 0,1446  | 15,32 | 0,000 |
| Efsolid   | 0,4357   | 0,1826  | 2,39  | 0,017 |
| HZFI      | -0,5732  | 0,1161  | -4,94 | 0,000 |
| HZEX      | -0,7049  | 0,3812  | -1,85 | 0,065 |
| NSUS      | -0,01120 | 0,08409 | -0,13 | 0,894 |
| TACCGAC   | 0,5992   | 0,1301  | 4,61  | 0,000 |
| CGEXT     | 0,3496   | 0,1312  | 2,67  | 0,008 |
| Menviade  | 1,0573   | 0,2080  | 5,08  | 0,000 |
| ns*hzex   | 0,7621   | 0,2564  | 2,97  | 0,003 |

S = 1,473      R-Sq = 12,9%      R-Sq(adj) = 12,1%

Analysis of Variance

| Source         | DF  | SS       | MS     | F     | P     |
|----------------|-----|----------|--------|-------|-------|
| Regression     | 8   | 297,068  | 37,133 | 17,11 | 0,000 |
| Residual Error | 926 | 2010,232 | 2,171  |       |       |
| Total          | 934 | 2307,299 |        |       |       |

Quadre 11: Llistat de Minitab del model final pel LNINJURED pels accidents amb ferits i amb la variable CAUSA GENERAL.

El model final per la variable **INJURED** és l'exponencial del model del Quadre 11 menys 0,1 (ja que per evitar la indeterminació del  $\ln(0)$ , es va sumar 0,1 a la variable **INJURED**).

Tot i que la variable **NSUS** i **HZEX** no són significativament diferent de zero, s'han deixat en el model final perquè la interacció **NS\*HZEX** és significativa.

A la Figura 65 s'observa el gràfic de normalitat dels residus del model final pel logaritme neperià de la variable **INJURED**. Els residus no segueixen una distribució del tot normal.

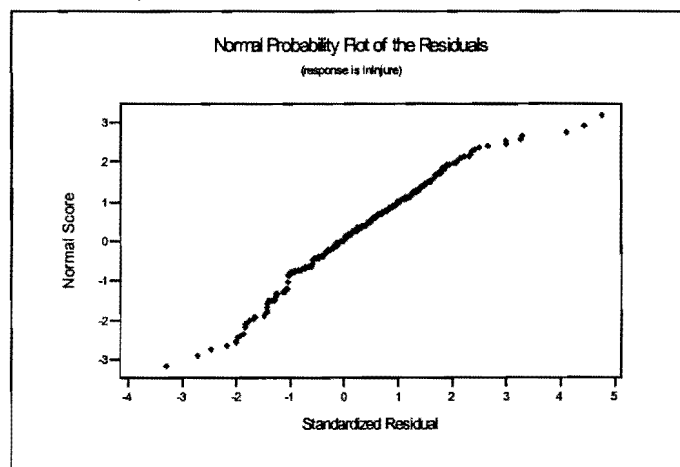


Figura 65: Gràfic de normalitat dels residus del model final del **LNINJURED** pels accidents amb ferits i amb la variable **CAUSA GENERAL**.

A la Figura 66 es poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de Injured, i els mateixos gràfics pel nombre de substàncies. En aquests s'hi detecten algunes anomalies, ja detectades abans, i queda clar el fet que els residus no acaben de ser aleatoris, ja que la variable **INJURED** és discreta. En els gràfics dels valors previstos es diferencien els accident amb un ferit de la resta, ja que hi ha molts accidents que tenen un ferit.



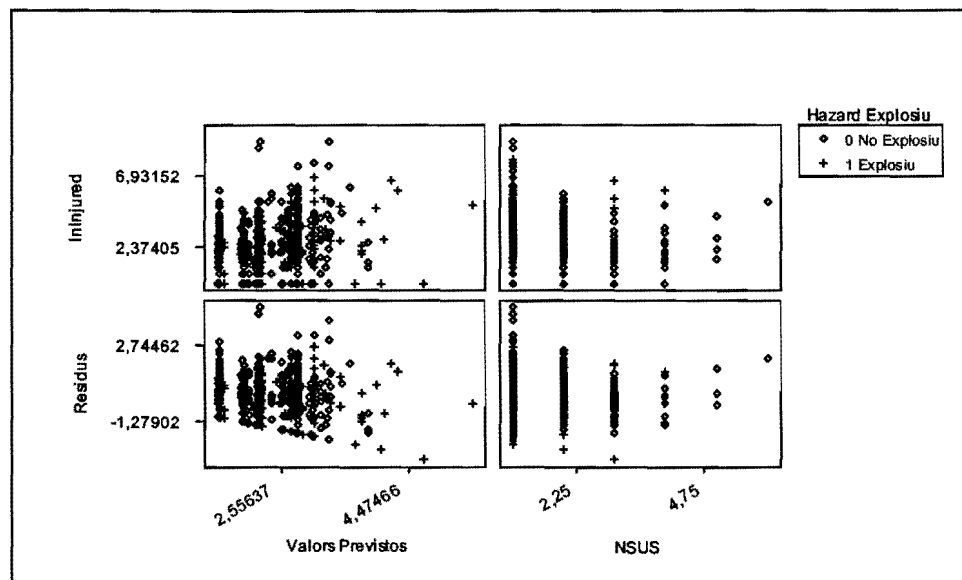


Figura 66: Draftman Plot del model final del LNINJURED pels accidents amb ferits i amb la variable CAUSA GENERAL.

## 7.2.2. Model sense la variable CAUSA GENERAL.

El model final al qual s'arriba amb les variables estat físic (EF\*), risc de la substància (HZ\*), nombre de substàncies (NSUS), tipus d'accident (TACC\*), tipus d'activitat (TACT\*), nivell de desenvolupament de l'estat (M\*) i els productes del nombre de substàncies amb la resta de variables indicadores (NS.\*) té les següents observacions amb un residu més gran que 3 en valor absolut:

| Obs  | HZTO | lnInjure | Fit    | StDev Fit | Residual | St Resid |
|------|------|----------|--------|-----------|----------|----------|
| 311  | 0,00 | 0,0953   | 4,2741 | 0,3645    | -4,1788  | -3,18RX  |
| 1987 | 0,00 | 5,8582   | 1,7323 | 0,0814    | 4,1259   | 3,03R    |
| 2004 | 0,00 | 6,9079   | 1,9928 | 0,0993    | 4,9150   | 3,62R    |
| 2005 | 0,00 | 6,9079   | 2,8359 | 0,0998    | 4,0719   | 3,00R    |
| 2007 | 0,00 | 7,3133   | 2,1489 | 0,1156    | 5,1644   | 3,80R    |
| 2008 | 1,00 | 7,6010   | 3,0960 | 0,1133    | 4,5050   | 3,32R    |
| 2009 | 0,00 | 7,6010   | 2,2661 | 0,1424    | 5,3348   | 3,94R    |
| 2010 | 0,00 | 7,8241   | 2,3332 | 0,0810    | 5,4909   | 4,04R    |
| 2011 | 0,00 | 8,0064   | 2,8359 | 0,0998    | 5,1705   | 3,80R    |
| 2012 | 1,00 | 8,6995   | 2,2529 | 0,0873    | 6,4467   | 4,74R    |
| 2013 | 1,00 | 9,2103   | 3,2945 | 0,1277    | 5,9158   | 4,36R    |
| 2014 | 1,00 | 9,2103   | 3,0960 | 0,1133    | 6,1143   | 4,50R    |

Quadre 12: Llistat d'anomalies del model pel LNINJURED pels accidents amb ferits i sense la variable CAUSA GENERAL.

Aquestes anomalies es caracteritzen una altra vegada per tenir un nombre elevat de ferits. Aquestes són:

- Les observacions 311, 2.009, 2.010, 2.012, 2.013 i 2.014 s'han comentat en l'apartat anterior.
- L'observació 1.987 té com codi, AN, el 2805, amb 350 ferits, aquest accident destaca per tenir un major nombre de ferits ja que segons el model trobat li corresponen 5,56 ferits i en té molts més.
- L'observació 2.004 té com codi, AN, el 2006, amb 1.000 ferits. Aquest accident es caracteritza per tenir un major nombre de ferits en relació als accidents produïts per una explosió.
- L'observació 2.005 té com codi, AN, el 2954, amb 1.000 ferits, aquest accident destaca per tenir un major nombre de ferits ja que segons el model trobat li corresponen 17,12 ferits i en té molts més.
- L'observació 2.007 té com codi, AN, el 8418, amb 1.500 ferits, aquest accident destaca per tenir un major nombre de ferits en relació als accidents produïts en locals domèstics o comercials (variable que pertany al model final).
- L'observació 2.008 té com codi, AN, el 2072, amb 2.000 ferits. Aquest accident ja havia estat detectat en el capítol 4 per tenir un nombre elevat de ferits en relació als accidents produïts en estats en via de desenvolupament.
- L'observació 2.011 té com codi, AN, el 2091, amb 3.000 ferits, aquest accident destaca per tenir un major nombre de ferits en relació als accidents produïts per una explosió amb substàncies explosives (variables que pertanyen al model final).

El model final pels accidents amb ferris i sense la causa general es pot observar al Quadre 13.

| $\ln \text{Injured} = 2,73 + 0,291 \text{ HZTO} - 0,950 \text{ HZFI} - 0,716 \text{ HZEX} -$ $- 0,389 \text{ HZCO} - 0,213 \text{ NSUS} + 0,315 \text{ TACCEX} +$ $+ 0,693 \text{ TACCGAC} + 0,292 \text{ TACCREL} - 0,184 \text{ TACTDO/C} -$ $- 0,843 \text{ Mdesenv} + 0,455 \text{ ns*hzfi} + 0,724 \text{ ns*hzex}$ |          |         |        |       |
|--|----------|---------|--------|-------|
| 1902 cases used 112 cases contain missing values   |          |         |        |       |
| Predictor  | Coef     | StDev   | T      | P     |
| Constant   | 2,7258   | 0,1795  | 15,19  | 0,000 |
| HZTO   | 0,29101  | 0,08325 | 3,50   | 0,000 |
| HZFI   | -0,9501  | 0,1754  | -5,42  | 0,000 |
| HZEX   | -0,7163  | 0,2577  | -2,78  | 0,006 |
| HZCO   | -0,3891  | 0,1179  | -3,30  | 0,001 |
| NSUS   | -0,2130  | 0,1254  | -1,70  | 0,090 |
| TACCEX   | 0,31519  | 0,08528 | 3,70   | 0,000 |
| TACCGAC  | 0,6933   | 0,1018  | 6,81   | 0,000 |
| TACCREL  | 0,29214  | 0,08880 | 3,29   | 0,001 |
| TACTDO/C   | -0,18431 | 0,08855 | -2,08  | 0,038 |
| Mdesenv  | -0,84313 | 0,08072 | -10,45 | 0,000 |
| ns*hzfi  | 0,4553   | 0,1379  | 3,30   | 0,001 |
| ns*hzex  | 0,7242   | 0,1959  | 3,70   | 0,000 |
| S = 1,363      R-Sq = 15,1%      R-Sq(adj) = 14,5%   |          |         |        |       |

Quadre 13: Llistat de Minitab del model final pel **LNINJURED** pels accidents amb ferits i sense la variable **CAUSA GENERAL**.

Els residus segueixen una distribució semblant a la Figura 65, que correspon al gràfic de normalitat dels residus del model final per tots els ferits amb la causa general. I a la Figura 67 es poden observar diferents gràfics dels valors previstos vs residus i el logaritme neperià de Injured, i els mateixos gràfics pel nombre de substàncies. Es tornen a detectar algunes anomalies, s'observa com els residus no són del tot aleatoris i també com es diferencien els accident amb un ferit de la resta, ja que hi ha molts accidents que tenen un ferit.

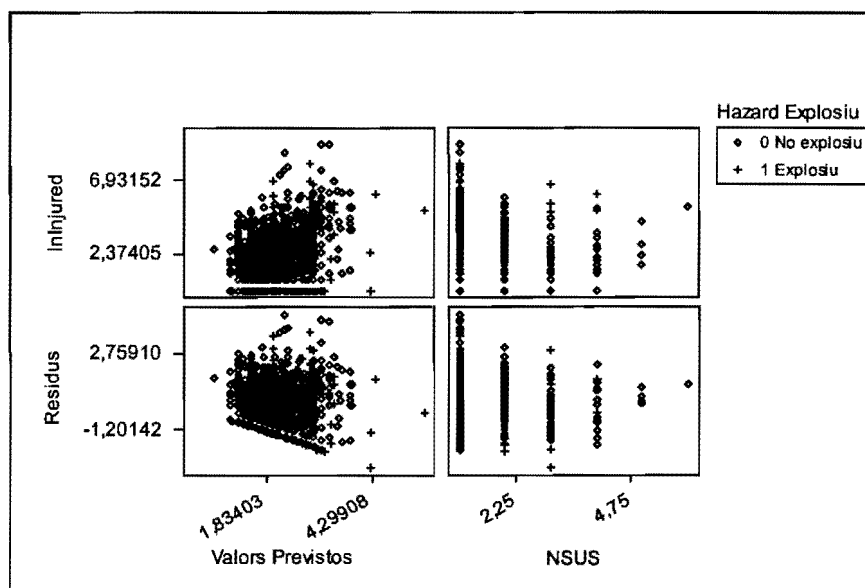


Figura 67: Draftman Plot del model final del LNINJURED pels accidents amb ferits i sense la variable CAUSA GENERAL.

## **7.3. Interpretació dels models pels accidents amb ferits**

A continuació s'interpreten els models trobats pels accidents amb ferits, tenint en compte la variable **CAUSA GENERAL** i sense tenir-la en compte.

### **7.3.1. Model amb la variable CAUSA GENERAL**

El model final pels accidents amb ferits fent servir la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Injured} = & 2,22 + 0,436 \text{ Efsolid} - 0,573 \text{ HZFI} - 0,705 \text{ HZEX} - \\ & - 0,0112 \text{ NSUS} + 0,599 \text{ TACCGAC} + 0,350 \text{ CGEXT} + \\ & + 1,06 \text{ Menviadesenv} + 0,762 \text{ ns*hzex} \end{aligned}$$

Quan en l'accident intervenen substàncies en estat físic sòlid, el nombre de ferits augmenta. Però si la substància o substàncies són inflamables el nombre de ferits disminueix. El nombre de ferits augmenta en 0,0458 per cada substància explosiva que intervingui en l'accident. Si l'accident és un núvol de gas el nombre de ferits augmenta. I el mateix passa quan la causa de l'accident és externa i quan l'accident es produeix en estats en via de desenvolupament.

En analitzar les variables explicatives que formen part del model final no sembla lògic, des del punt de vista de l'enginyer químic responsable del projecte, els signes negatius dels coeficients de les variables **HZFI** i **HZEX**. S'observa que existeix dependència entre les variables **HZEX** i **EFSSOLID**. Ja que més del 77% de les substàncies explosives estan en estat físic sòlid. La variable **HZFI** està correlacionada amb la variable **CGEXT**, ja que en més de 65% dels accidents produïts per una causa externa han participat substàncies inflamables.

La variable **MDESENV** està molt correlacionada amb totes les variables explicatives ja que més del 78% dels accidents de la base de dades s'han produït en estats desenvolupats.

Les variables regressores expliquen un 12,1% el nombre de ferits en accidents industrials amb substàncies perilloses, això pot ser degut a que hi ha altres variables que expliquen el nombre de ferits, però que no hi són a la base de dades MHIDAS.

### 7.3.2. Model sense la variable CAUSA GENERAL

El model final pels accidents amb ferits sense fer servir la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Injured} = & 2,73 + 0,291 \text{ HZTO} - 0,950 \text{ HZFI} - 0,716 \text{ HZEX} - \\ & - 0,389 \text{ HZCO} - 0,213 \text{ NSUS} + 0,315 \text{ TACCEX} + \\ & + 0,693 \text{ TACCGAC} + 0,292 \text{ TACCREL} - 0,184 \text{ TACTDO/C} - \\ & - 0,843 \text{ Mdesenv} + 0,455 \text{ ns*hzfi} + 0,724 \text{ ns*hzex} \end{aligned}$$

Quan en l'accident intervenen substàncies tòxiques el nombre de ferits augmenta. Si les substàncies són corrosives el nombre de ferits en canvi disminueix. El nombre de ferits augmenta en 0,242 per cada substància inflamable que intervingui en l'accident i 0,511 per cada substància explosiva. Si l'accident és una explosió el nombre de ferits també augmenta. El mateix passa si és un núvol de gas o una fuga. Per contra el nombre de ferits disminueix si l'accident es produeix en locals domèstics o comercials i si es produeix en estats desenvolupats.

En analitzar les variables explicatives que formen part del model final no sembla lògic, des del punt de vista de l'enginyer químic responsable del projecte, els signes negatius dels coeficients de les variables **HZFI**, **HZEX** i **HZCO**.

S'observa que les variables **HZEX** i **TACCEX** estan correlacionades, ja que el 92% de substàncies explosives intervenen en accidents produïts per una explosió.

La variable **TACCEX** també està correlacionada amb la variable **HZFI**, ja que en més del 64% dels accidents produïts per una explosió participen substàncies inflamables.

Les variables **HZTO** i **TACCGAC** estan correlacionades, ja que en més del 84% dels accidents produïts per un núvol de gas participen substàncies tòxiques. La variable **HZTO** també està correlacionada amb la variable **HZCO**, ja que en més del 50% de les substàncies corrosives són tòxiques.

La variable **TACCREL** està correlacionada amb les variables **HZTO**, **HZCO**, **TACCGAC** i **TACTDO/C**. Ja que el 65% de les substàncies tòxiques i un 71% de les substàncies corrosives han participat en accidents produïts per una fuga. A més a més, en un 65% dels accidents produïts per un núvol de gas també s'ha produït una fuga. I en un 50% dels accidents produïts en locals domèstics o comercials s'han originat per una fuga.

També cal tenir present les correlacions mencionades en l'apartat anterior. A més a més, les variables que estan correlacionades s'han d'interpretar conjuntament.

Les variables regressores expliquen un 14,5% el nombre de ferits en accidents industrials amb substàncies perilloses.

## **7.4. Models per tots els accidents (inclosos els sense ferits)**

Per construir el model per tots els accidents tant pels que tenen ferits com pels que no en tenen, s'ha utilitzat la base de dades amb 2.793 accidents. La variable **QSUS** té 2.156 missings i com aquest fet redueix dràsticament el nombre de graus de llibertat que es disposen no es farà servir, ni el logaritme neperià de la quantitat de substància (**LNQSUS**) ni els productes de **lnqsus** per les diferents variables indicadores. Les causes generals (**CG\***) també presenten un nombre important de missings. S'ha decidit entrar-la primer en el model encara que redueixi gairebé a la meitat els graus de llibertat disponibles, i després construir un altre model sense tenir en compte aquesta variable. Aquest últim model tindrà un 46% més d'observacions i això farà que el model final sigui un altre.

| VARIABLE                          | MISSINGS |
|-----------------------------------|----------|
| ESTAT FÍSIC<br>(EF*)              | 250      |
| RISC DE LA SUBSTÀNCIA<br>(HZ*)    | 0        |
| NOMBRE DE SUBSTÀNCIES<br>(NSUS)   | 0        |
| TIPUS D'ACCIDENT<br>(TACC*)       | 79       |
| TIPUS D'ACTIVITAT<br>(TACT*)      | 79       |
| CAUSA GENERAL<br>(CG*)            | 1.282    |
| NIVELL DE DESENVOLUPAMENT<br>(M*) | 0        |
| QSUS                              | 2.156    |

Taula 37: Nombre de missings de les variables explicatives quan es modela per tots els accidents. El nombre total d'accidents és 2.793.



### 7.4.1. Model amb la variable CAUSA GENERAL.

El model final al qual s'arriba amb la variable **CAUSA GENERAL** té les següents observacions amb un residu més gran que 3, en valor absolut:

| Obs  | Eflíquid | lnInjure | Fit    | StDev Fit | Residual | St Resid |
|------|----------|----------|--------|-----------|----------|----------|
| 2788 | 1,00     | 7,6010   | 0,5833 | 0,2545    | 7,0177   | 3,13R    |
| 2791 | 1,00     | 8,6995   | 0,7167 | 0,2069    | 7,9828   | 3,55R    |
| 2792 | 1,00     | 9,2103   | 2,0966 | 0,2977    | 7,1136   | 3,18R    |

Quadre 14: Llistat d'anomalies del model final per **LNINJURED** per tots els accidents i amb la variable **CAUSA GENERAL**.

Aquestes anomalies tornen a ser els accidents amb un nombre elevat de ferits. Aquestes són:

- Les observacions 2.788 (AN 6324), 2.791 (AN 810) i 2.792 (AN 1098), s'han comentat en apartats anteriors.

El model final per la variable **INJURED** és l'exponencial del model del Quadre 15 menys 0,1.

| $\ln \text{Injured} = 0,931 - 0,425 \text{ Eflíquid} + 0,540 \text{ HZTO} - 0,813 \text{ HZFI} +$ $+ 0,266 \text{ NSUS} + 1,15 \text{ TACCEX} + 0,521 \text{ TACCGAC} +$ $+ 0,518 \text{ TACTDO/C} + 0,413 \text{ TACTPRO} + 0,550 \text{ TACTTRANS} +$ $+ 0,664 \text{ CGHUM} - 1,26 \text{ Mdesenv}$ |         |         |        |       |       |
|--|---------|---------|--------|-------|-------|
| 1383 cases used 1410 cases contain missing values  |         |         |        |       |       |
| Predictor  | Coef    | StDev   | T      | P     |       |
| Constant   | 0,9309  | 0,2646  | 3,52   | 0,000 |       |
| Eflíquid   | -0,4249 | 0,1301  | -3,27  | 0,001 |       |
| HZTO   | 0,5396  | 0,1484  | 3,64   | 0,000 |       |
| HZFI   | -0,8130 | 0,1410  | -5,76  | 0,000 |       |
| NSUS   | 0,2659  | 0,1066  | 2,49   | 0,013 |       |
| TACCEX   | 1,1515  | 0,1406  | 8,19   | 0,000 |       |
| TACCGAC  | 0,5207  | 0,1773  | 2,94   | 0,003 |       |
| TACTDO/C   | 0,5178  | 0,2235  | 2,32   | 0,021 |       |
| TACTPRO  | 0,4130  | 0,1433  | 2,88   | 0,004 |       |
| TACTTRAN   | 0,5498  | 0,1834  | 3,00   | 0,003 |       |
| CGHUM  | 0,6643  | 0,1268  | 5,24   | 0,000 |       |
| Mdesenv  | -1,2592 | 0,1866  | -6,75  | 0,000 |       |
| S = 2,258      R-Sq = 16,8%      R-Sq(adj) = 16,2%   |         |         |        |       |       |
| Analysis of Variance   |         |         |        |       |       |
| Source   | DF      | SS      | MS     | F     | P     |
| Regression   | 11      | 1416,14 | 128,74 | 25,24 | 0,000 |
| Residual Error   | 1371    | 6992,60 | 5,10   |       |       |
| Total  | 1382    | 8408,73 |        |       |       |

Quadre 15: Llistat de Minitab del model final per **LNINJURED** per tots els accidents i amb la variable **CAUSA GENERAL**.

A la Figura 68 s'observa el gràfic de normalitat dels residus del model final pel logaritme neperià de la variable **INJURED**. I a la Figura 69 s'hi poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de Injured, i els mateixos gràfics pel nombre de substàncies. En ambdós gràfics es detecten anomalies. També s'observa com els residus no acaben de ser aleatoris. Els accident que no tenen cap ferit es diferencien de la resta molt clarament, ja que hi ha molts accidents que no tenen cap ferit.

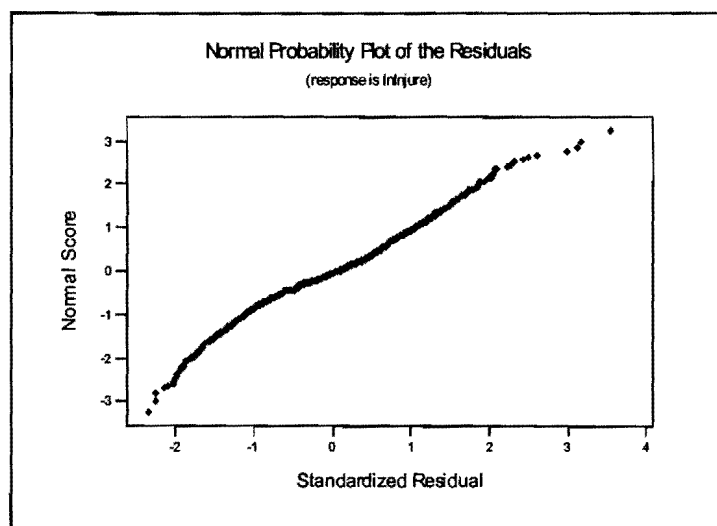


Figura 68: Gràfic de normalitat dels residus del model final del **LNINJURED** per tots els accidents i amb la variable **CAUSA GENERAL**.

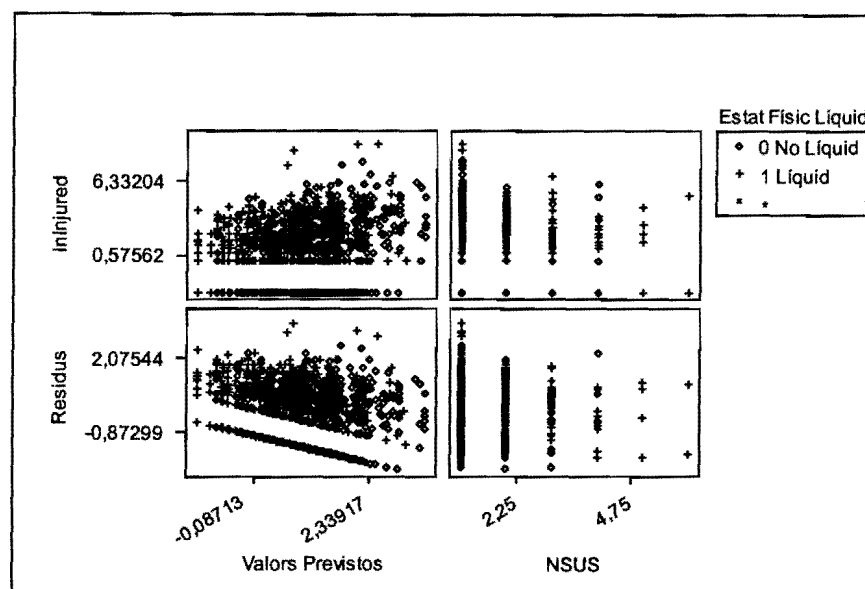


Figura 69: Draftman Plot del model final pel **LNINJURED** per tots els accidents i amb la variable **CAUSA GENERAL**.

## 7.4.2. Model sense la variable CAUSA GENERAL.

El model final al qual s'arriba sense la variable **CAUSA GENERAL** té els següents valors anòmals:

| Obs  | Eflíquid | lnInjure | Fit    | StDev Fit | Residual | St Resid |
|------|----------|----------|--------|-----------|----------|----------|
| 2788 | 1,00     | 7,6010   | 0,5212 | 0,2110    | 7,0798   | 3,33R    |
| 2791 | 1,00     | 8,6995   | 0,4835 | 0,1478    | 8,2160   | 3,85R    |
| 2792 | 1,00     | 9,21     | 1,5767 | 0,2137    | 7,6335   | 3,59R    |
| 2793 | 1,00     | 9,2103   | 2,4851 | 0,2511    | 6,7252   | 3,17R    |

Quadre 16: Llistat d'anomalies del model final per **LNINJURED** per tots els accidents i sense la variable **CAUSA GENERAL**.

Aquestes anomalies es caracteritzen per tenir un nombre elevat de ferits. Són:

- Les observacions 2.788 (AN 6324), 2.791 (AN 810), 2.792 (AN 1098) i 2.793 (AN 2744) s'han comentat en apartats anteriors.

| $\text{lnInjured} = 1,06 - 0,369 \text{ Eflíquid} + 0,555 \text{ HZTO} - 0,799 \text{ HZFI} -$ $- 0,454 \text{ HZCO} + 0,259 \text{ NSUS} + 1,00 \text{ TACCEX} +$ $+ 0,547 \text{ TACCGAC} + 0,511 \text{ TACTDO/C} + 0,320 \text{ TACTPRO} +$ $+ 0,500 \text{ TACTTRANS} - 1,03 \text{ Mdesenv} + 0,476 \text{ Menviadesenv}$ |          |         |       |       |
|---|----------|---------|-------|-------|
| 2419 cases used 374 cases contain missing values  |          |         |       |       |
| Predictor   | Coef     | StDev   | T     | P     |
| Constant  | 1,0641   | 0,1900  | 5,60  | 0,000 |
| Eflíquid  | -0,36910 | 0,09701 | -3,80 | 0,000 |
| HZTO  | 0,5551   | 0,1100  | 5,04  | 0,000 |
| HZFI  | -0,7994  | 0,1084  | -7,38 | 0,000 |
| HZCO  | -0,4541  | 0,1645  | -2,76 | 0,006 |
| NSUS  | 0,25922  | 0,08411 | 3,08  | 0,002 |
| TACCEX  | 1,00026  | 0,09964 | 10,04 | 0,000 |
| TACCGAC   | 0,5468   | 0,1409  | 3,88  | 0,000 |
| TACTDO/C  | 0,5115   | 0,1469  | 3,48  | 0,001 |
| TACTPRO   | 0,3201   | 0,1028  | 3,11  | 0,002 |
| TACTTRAN  | 0,5001   | 0,1423  | 3,52  | 0,000 |
| Mdesenv   | -1,0258  | 0,1397  | -7,34 | 0,000 |
| Menviade  | 0,4757   | 0,2157  | 2,21  | 0,028 |
| S = 2,139      R-Sq = 16,6%      R-Sq(adj) = 16,2%  |          |         |       |       |

Quadre 17: Llistat de Minitab del model final per **LNINJURED** per tots els accidents i sense la variable **CAUSA GENERAL**.

El model final per la variable **INJURED** és l'exponencial del model del Quadre 17 menys 0,1.

La distribució dels residus es semblant a la vista en l'apartat anterior. A la Figura 70 s'hi poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de Injured, i els mateixos gràfics pel nombre de substàncies. Es detecta alguna anomalia i com els residus no acaben de ser aleatoris. I com els accident amb cap ferit es diferencien de la resta molt clarament i els que tenen un ferit també, però no tant.

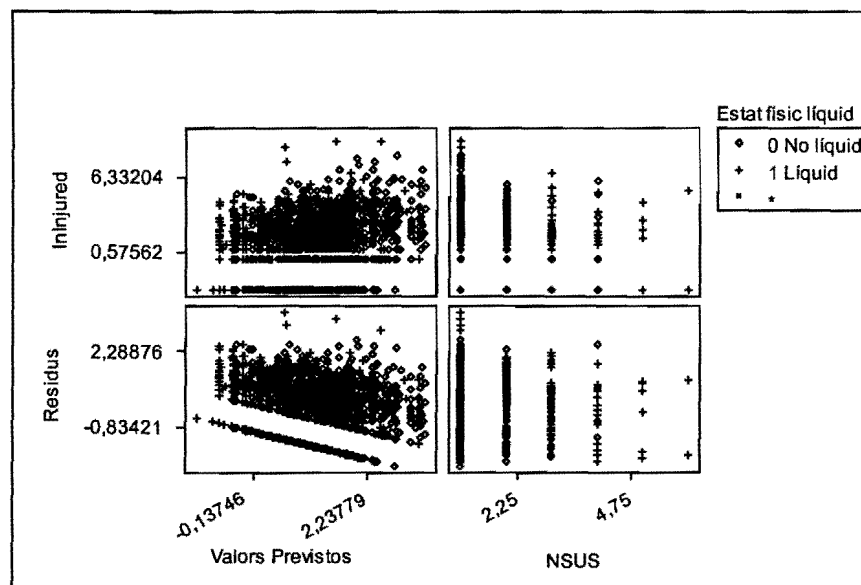


Figura 70: Draftman Plot del model final pel **LNINJURED** per tots els accidents i sense la variable **CAUSA GENERAL**.

## **7.5. Interpretació dels models per tots els accidents (inclosos els sense ferits)**

A continuació s'interpreten els models trobats per tots els accidents, tant pels accidents que tenen ferits com pels que no en tenen, tenint en compte la variable **CAUSA GENERAL** i sense tenir-la en compte.

### **7.5.1. Model amb la variable CAUSA GENERAL**

El model final per tots els accidents fent servir la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Injured} = & 0,931 - 0,425 \text{ Eflíquid} + 0,540 \text{ HZTO} - 0,813 \text{ HZFI} + \\ & + 0,266 \text{ NSUS} + 1,15 \text{ TACCEX} + 0,521 \text{ TACCGAC} + \\ & + 0,518 \text{ TACTDO/C} + 0,413 \text{ TACTPRO} + 0,550 \text{ TACTTRANS} + \\ & + 0,664 \text{ CGHUM} - 1,26 \text{ Mdesenv} \end{aligned}$$

Quan en l'accident intervenen substàncies en estat físic líquid, el nombre de ferits disminueix. Però si la substància o substàncies són tòxiques el nombre de ferits augmenta. El nombre de ferits disminueix si la substància és inflamable. En canvi el nombre de ferits augmenta en augmentar el nombre de substàncies, si l'accident és una explosió, el mateix passa si és un núvol de gas, si l'accident es produeix en locals domèstics o comercials, o en una planta de procés, o durant operacions de càrrega o descàrrega. També el nombre de ferits augmenta quan l'accident es produeix per causa humana. En canvi, quan l'accident es produeix en estats desenvolupats el nombre de ferits disminueix.

En analitzar les variables explicatives que formen part del model final es conclou que la variable **HZFI** està correlacionada amb les variables **EFLÍQUID**, **TACCEX**, **TACTPRO**, **TACTTRANS** i **CGHUM**. Ja que el 74% de les substàncies en estat físic líquid són inflamables, en més del 67% dels accidents produïts per una explosió participen substàncies inflamables. En el 65% dels accidents produïts en plantes de procés, en el 71% produïts durant operacions de càrrega o descàrrega i en un 64% produïts per una errada humana, han participat substàncies inflamables.

Les variables **TACCGAC** i **HZTO** estan correlacionades, ja que en més del 79% dels accidents produïts per un núvol de gas intervenen substàncies tòxiques.

Les variables **TACTPRO** i **TACCEX** estan correlacionades, ja que més del 51% dels accidents produïts en plantes de procés es produeixen per una explosió.

També cal recordar la correlació de la variable **MDESENV** amb la resta de variables explicatives. El 82,2% dels accidents on es coneix el nombre de ferits es produeixen en estats desenvolupats.

Les variables regressores expliquen un 16,2% el nombre de ferits en accidents industrials amb substàncies perilloses.

## 7.5.2. Model sense la variable CAUSA GENERAL

El model final per tots els accidents amb ferits sense fer servir la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Injured} = & 1,06 - 0,369 \text{ Eflíquid} + 0,555 \text{ HZTO} - 0,799 \text{ HZFI} - \\ & - 0,454 \text{ HZCO} + 0,259 \text{ NSUS} + 1,00 \text{ TACCEX} + \\ & + 0,547 \text{ TACCGAC} + 0,511 \text{ TACTDO/C} + 0,320 \text{ TACTPRO} + \\ & + 0,500 \text{ TACTTRANS} - 1,03 \text{ Mdesenv} + 0,476 \text{ Menviadesenv} \end{aligned}$$

Quan en l'accident intervenen substàncies en estat físic líquid el nombre de ferits disminueix. Si la substància o substàncies són tòxiques el nombre de ferits augmenta. En canvi, si les substàncies són inflamables el nombre de ferits disminueix. El mateix passa si són corrosives. El nombre de ferits augmenta en augmentar el nombre de substàncies, si l'accident és una explosió, el mateix passa si és un núvol de gas, si l'accident s'ha produït en locals domèstics o comercials, o en una planta de procés, o durant operacions de càrrega o descàrrega. D'altra banda, el nombre de ferits disminueix si l'accident es produeix en estats desenvolupats. En canvi augmenta si es produeix en estats en via de desenvolupament.

A més a més, de les correlacions mencionades en l'apartat anterior cal destacar la correlació entre la variable **HZCO** i **EFLÍQUID**. Un 80,14% de les substàncies corrosives es troben en estat físic líquid.

Les variables regressores expliquen un 16,2% el nombre de ferits en accidents industrials amb substàncies perilloses.



## **7.6. Conclusions**

En analitzar els resultats dels diferents models, es pot concloure que els models pels que no tenen ferits i per tots els accidents (inclosos els que no en tenen) són diferents, tant si es té en compte la variable **CAUSA GENERAL** com si no. Quan es torna a ajustar el model sense tenir en compte la variable **CAUSA GENERAL** hi ha variables que deixen de ser significativament diferent de zero i altres que passen a ser-ho. Cal destacar que en no tenir en compte la variable **CAUSA GENERAL** pels accidents amb ferits el model explica més i té menys variabilitat. En canvi per tots els accidents el model explica el mateix però la variabilitat disminueix. A més a més, els models amb ferits tenen menys variabilitat en relació als models per a tots els accidents.

En els quatre models ajustats sempre hi són les variables **HZFI**, **NSUS** i **TACCGAC**. La primera variable sempre és negativa, però això no és lògic des del punt de vista de l'enginyer químic responsable del projecte. Pot ser degut a la colinealitat que té aquesta variable amb altres explicatives. Pel que fa a la variable **NSUS** en els models pels accidents amb apareixen interaccions d'aquesta variable amb **HZFI** i **HZEX**. En canvi en els models per tots els accidents no apareixen interaccions. Per últim la variable **TACCGAC** sempre té un coeficient positiu, el que significa que quan l'accident es produeix per un núvol de gas el nombre de ferits augmenta.

Els models trobats per a **INJURED** que més expliquen són per tots els accidents. Però si no es té en compte la variable **CAUSA GENERAL** la variabilitat disminueix. Les variables regressores expliquen un 16,2% la variable logaritme neperià del nombre de ferits.

També cal destacar que hi ha colinealitat entre algunes variables explicatives. I en aquesta situació les variables s'han d'explicar conjuntament. A més a més, el model seria més explicatiu que predictiu, per les raons comentades en el capítol anterior.

# CAPÍTOL 8: MODEL LINEAL PER A DAMAGE

## 8.1. Introducció

En aquest capítol es construïran models lineals pel logaritme neperià de l'estimació econòmica del dany material en accidents industrials amb substàncies perilloses, és a dir, per a la variable **LNDAMAGE**.

També cal recordar que hi ha variables que contenen un gran nombre de missings. La variable **DAMAGE** presenta 4.417 missings. Tal i com s'ha fet per les anteriors variables resposta, s'ha creat una nova base de dades eliminant els accidents en els que es desconeix l'estimació econòmica del dany material. La nova base de dades conté un total de 750 accidents. Un cop feta aquesta selecció, cal veure quines variables explicatives tenen prou observacions i quines no.

La variable **QSUS**, tal i com s'observa a la Taula 38, té 542 missings, i com aquest fet redueix dràsticament el nombre de graus de llibertat disponibles no entrarà en el model, ni el logaritme neperià de la quantitat de substància (**LNQSUS**) ni els productes d'aquesta variable per les diferents variables indicadores. Les causes generals (**CG\***) també presenten un nombre important de missings, 265. S'ha decidit entrar-la en el model encara que redueix gairebé a la meitat els graus de llibertat disponibles, però si en el model final hi ha alguna causa general aleshores es construirà un altre model sense tenir en compte aquesta variable. Aquest últim model tindrà un 35% més d'observacions i això pot fer que el model final sigui un altre.

A continuació es mostra el llistat dels missings que tenen les variables explicatives pels accidents on es coneix l'estimació econòmica del dany material.

| VARIABLE                        | MISSINGS |
|---------------------------------|----------|
| ESTAT FÍSIC<br>(EF*)            | 59       |
| RISC DE LA SUBSTÀNCIA<br>(HZ*)  | 0        |
| NOMBRE DE SUBSTÀNCIES<br>(NSUS) | 0        |
| TIPUS D'ACCIDENT<br>(TACC*)     | 15       |
| TIPUS D'ACTIVITAT<br>(TACT*)    | 19       |
| CAUSA GENERAL<br>(CG*)          | 265      |
| NIVELL DE<br>DESENVOLUPAMENT    | 0        |
| QSUS                            | 542      |

Taula 38: Nombre de missings de les variables explicatives per tots els accidents, el nombre total és 750.

## 8.2. Model amb la variable CAUSA GENERAL

El model final al qual s'arriba amb les variables estat físic (EF\*), risc de la substància (HZ\*), nombre de substàncies (NSUS), tipus d'accident (TACC\*), tipus d'activitat (TACT\*), causa general (CG\*), nivell de desenvolupament de l'estat (M\*) i els productes del nombre de substàncies amb la resta de variables indicadores (NS·\*) no té anomalies amb un residu més gran que 3, en valor absolut.

El model final per la variable **DAMAGE** és l'exponencial del model següent menys 0,1.

|   |         |          |        |       |       |
|---|---------|----------|--------|-------|-------|
| $\ln \text{Damage} = 0,925 + 0,549 \text{ Efgas} - 0,856 \text{ HZOX} +$ $+ 0,615 \text{ TACCEX} + 0,529 \text{ TACCREL} -$ $- 1,29 \text{ TACTDO/C} + 0,387 \text{ CGEXT}$ |         |          |        |       |       |
| 455 cases used 295 cases contain missing values   |         |          |        |       |       |
| Predictor   | Coef    | StDev    | T      | P     |       |
| Constant  | 0,9250  | 0,1662   | 5,57   | 0,000 |       |
| Efgas   | 0,5493  | 0,2242   | 2,45   | 0,015 |       |
| HZOX  | -0,8560 | 0,4006   | -2,14  | 0,033 |       |
| TACCEX  | 0,6149  | 0,1784   | 3,45   | 0,001 |       |
| TACCREL   | 0,5286  | 0,1982   | 2,67   | 0,008 |       |
| TACTDO/C  | -1,2862 | 0,4311   | -2,98  | 0,003 |       |
| CGEXT   | 0,3871  | 0,1961   | 1,97   | 0,049 |       |
| S = 1,785                  R-Sq = 7,9%                  R-Sq(adj) = 6,6%  |         |          |        |       |       |
| Analysis of Variance  |         |          |        |       |       |
| Source  | DF      | SS       | MS     | F     | P     |
| Regression  | 6       | 122,128  | 20,355 | 6,39  | 0,000 |
| Residual Error  | 448     | 1427,806 | 3,187  |       |       |
| Total   | 454     | 1549,934 |        |       |       |

Quadre 18: Llistat de Minitab del model final pel **LNDAMAGE** per a tots els accidents i amb la variable **CAUSA GENERAL**.

A la figura 71 s'observa el gràfic de normalitat dels residus del model final pel logaritme neperià de la variable **DAMAGE**. S'observa com els residus segueixen una distribució normal i la presència d'algunes observacions atípiques.

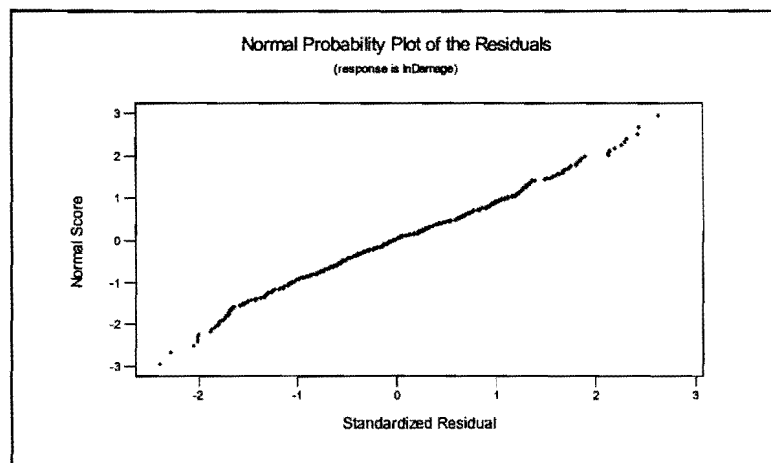


Figura 71: Normal Plot dels residus del model final **LNDAMAGE** per tots els accidents i amb la variable **CAUSA GENERAL**.

A la figura 72 es poden observar els gràfics dels valors previstos vs residus i el logaritme neperià de Damage, i els mateixos gràfics pel nombre de substàncies. En aquests s'hi detecten algunes anomalies, ja detectades abans, i s'observa que els residus són aleatoris, ja que la variable **DAMAGE** és continua.

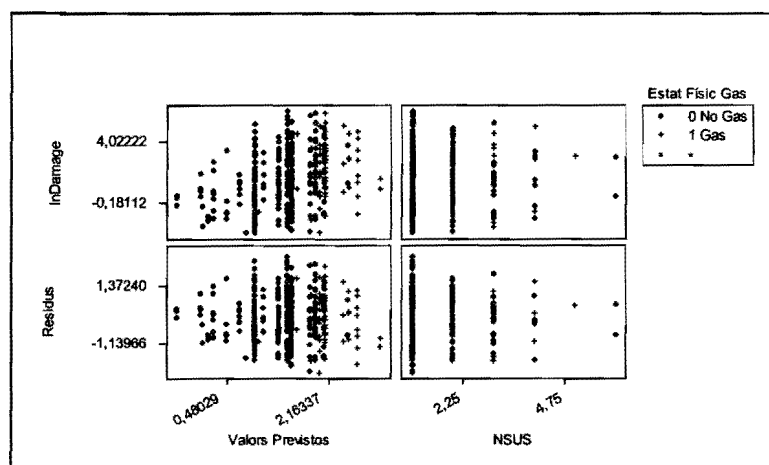


Figura 72: Draftman Plot del model final del **LNDAMAGE** per tots els accidents i amb la variable **CAUSA GENERAL**.

## **8.3. Interpretació del model amb la variable CAUSA GENERAL**

El model final per l'estimació econòmica del dany material per tots els accidents i incloent la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln\text{Damage} = & 0,925 + 0,549 \text{ Efgas} - 0,856 \text{ HZOX} + \\ & + 0,615 \text{ TACCEX} + 0,529 \text{ TACCREL} - \\ & - 1,29 \text{ TACTDO/C} + 0,387 \text{ CGEXT} \end{aligned}$$

L'estimació econòmica del dany material augmentarà si la substància o substàncies que intervenen en l'accident estan en estat físic gas. L'estimació econòmica disminueix si la substància és oxidant. En canvi augmenta si l'accident és una explosió o una fuga. Disminueix si l'accident es produeix en locals domèstics o comercials. I augmenta quan es produeix per una causa externa.

La variable **TACCEX** està correlacionada amb les variables **EFGAS**, **HZOX** i **TACTDO/C**. Ja que el 58,2% de les substàncies en estat físic gas i més del 64% de les substàncies oxidants han intervingut en una explosió. I el 52% dels accidents produïts en locals domèstics o comercials s'han produït per una explosió.

La variable **HZOX** i la variable **CGEXT** estan correlacionades, ja que més del 60% de les substàncies oxidants han intervingut en accidents que s'han produït per una causa externa.

Aquestes variables que tenen colinealitats s'han d'interpretar conjuntament a l'hora d'explicar com afecten a la variable resposta.

Per finalitzar la interpretació cal destacar que les variables regressores expliquen un 6,6% l'estimació econòmica de dany material en accidents industrials amb substàncies perilloses.

## 8.4. Model sense la variable CAUSA GENERAL.

El model final al qual s'arriba amb les variables mencionades abans i sense tenir en compte la variable **CAUSA GENERAL** no té cap anomalia amb un residu més gran que 3, en valor absolut.

El model final per la variable **DAMAGE** és l'exponencial del model següent menys 0,1.

| The regression equation is  |         |        |       |       |
|---|---------|--------|-------|-------|
| $\ln \text{Damage} = 0,926 + 0,624 \text{ Efgas} + 0,127 \text{ NSUS} + 0,308 \text{ TACCEX} +$ $+ 1,82 \text{ TACCGAC} + 0,424 \text{ TACCREL} - 0,976 \text{ TACTDO/C} -$ $- 1,59 \text{ TACTTRANS} - 1,57 \text{ ns*tacgc} + 1,02 \text{ ns*tatran}$ |         |        |       |       |
| 666 cases used 84 cases contain missing values  |         |        |       |       |
| Predictor   | Coef    | StDev  | T     | P     |
| Constant  | 0,9260  | 0,1833 | 5,05  | 0,000 |
| Efgas   | 0,6241  | 0,1869 | 3,34  | 0,001 |
| NSUS  | 0,1267  | 0,1090 | 1,16  | 0,246 |
| TACCEX  | 0,3078  | 0,1461 | 2,11  | 0,036 |
| TACCGAC   | 1,8242  | 0,6412 | 2,84  | 0,005 |
| TACCREL   | 0,4240  | 0,1697 | 2,50  | 0,013 |
| TACTDO/C  | -0,9756 | 0,2763 | -3,53 | 0,000 |
| TACTTRAN  | -1,5921 | 0,5736 | -2,78 | 0,006 |
| ns*tacgc  | -1,5665 | 0,4003 | -3,91 | 0,000 |
| ns*tatra  | 1,0222  | 0,4574 | 2,23  | 0,026 |
| S = 1,759      R-Sq = 8,8%      R-Sq(adj) = 7,6%  |         |        |       |       |

Quadre 19: Llistat de Minitab del model final del **LNDAMAGE** per tots els accidents i sense la variable **CAUSA GENERAL**.

A la figura 73 s'observa el gràfic de normalitat dels residus del model final pel logaritme neperià de la variable **DAMAGE**, aquests residus segueixen una distribució normal.

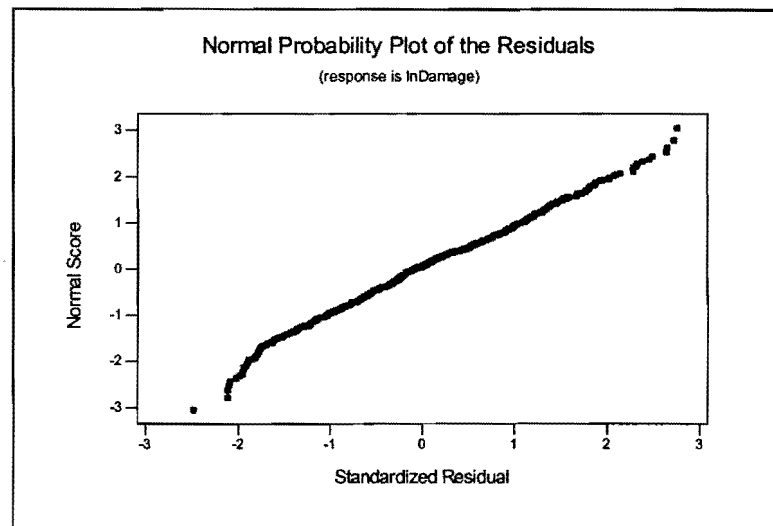


Figura 73: Normal Plot dels residus del model final del **LNDAMAGE** per tots els accidents i sense la variable **CAUSA GENERAL**.

En els gràfics dels valors previstos vs residus i logaritme neperià de Damage, i els mateixos gràfics pel nombre de substàncies (Figura 74) s'observa algunes anomalies, com també s'observaven en el gràfic anterior, i com els residus són aleatoris.

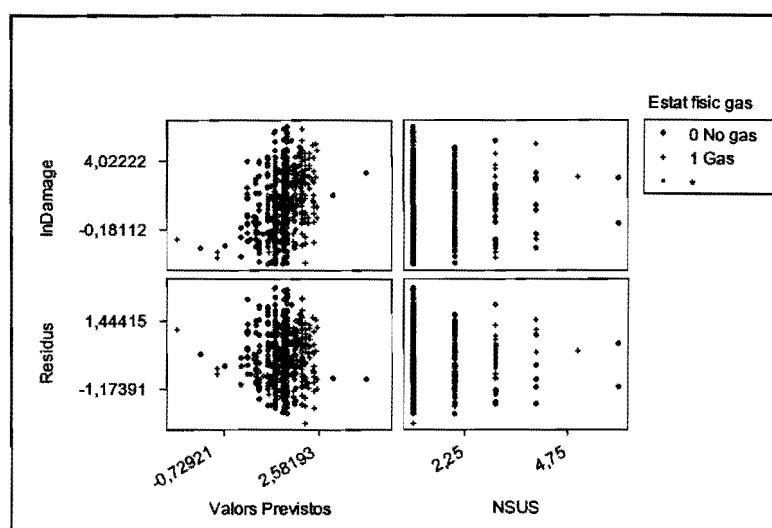


Figura 74: Draftman Plot del model final del **LNDAMAGE** per tots els accidents i sense la variable **CAUSA GENERAL**.



## **8.5. Interpretació del model sense la variable CAUSA GENERAL**

El model final per l'estimació econòmica del dany material per tots els accidents i sense la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Damage} = & 0,926 + 0,624 \text{ Efgas} + 0,127 \text{ NSUS} + 0,308 \text{ TACCEX} + \\ & + 1,82 \text{ TACCGAC} + 0,424 \text{ TACCREL} - 0,976 \text{ TACTDO/C} - \\ & - 1,59 \text{ TACTTRANS} - 1,57 \text{ ns*tacgc} + 1,02 \text{ ns*tatran} \end{aligned}$$

L'estimació econòmica del dany material augmenta si la substància o substàncies que intervenen en l'accident estan en estat físic gas i augmenta en 0,377 milions de dòlars per cada substància que intervingui en un accident produït per un núvol de gas. En canvi disminueix en 0,443 per cada substància que intervingui en un accident que es produeixi durant operacions de càrrega o descàrrega. L'estimació augmenta si l'accident es produeix per una explosió. I en canvi, disminueix si l'accident es produeix en locals domèstics o comercials.

A més a més de les correlacions comentades en l'apartat anterior, cal també destacar la correlació entre les variables **TACTTRANS** i **TACCEX**. En un 57,58% dels accidents que es produeixen durant operacions de càrrega i descàrrega s'ha produït una explosió.

Les variables regressores expliquen un 7,6% l'estimació econòmica del dany material en accidents industrials amb substàncies perilloses.

## 8.6. Conclusions

En analitzar els resultats dels diferents models, es pot concloure que el model amb la variable **CAUSA GENERAL** és diferent respecte al que no la inclou. Quan es torna a ajustar el model sense tenir en compte la variable **CAUSA GENERAL** hi ha variables que deixen de ser significativament diferent de zero i altres que passen a ser-ho. Cal destacar que en no tenir en compte la variable **CAUSA GENERAL** el model explica més i té menys variabilitat.

En els dos models ajustats sempre hi són les variables **EFGAS**, **TACCEX**, **TACCREL** i **TACTDO/C**. La primera variable pels dos model té un coeficient positiu, el que significa que quan en l'accident participen substàncies en estat físic gas l'estimació econòmica del dany material augmenta. El mateix succeeix quan l'accident es produeix per una explosió i/o per una fuga. En canvi, quan l'accident es produeix en locals domèstics o comercials disminueix.

El model trobat per a **DAMAGE** que més explica i menys variabilitat té és el que ajusta tots els accidents sense tenir en compte la variable **CAUSA GENERAL**, amb una  $R_{adj} = 7,6\%$ .

Cal destacar que hi ha colinealitat entre algunes variables explicatives i en aquest cas les variables correlacionades s'han d'interpretar conjuntament. I que el model seria més explicatiu que predictiu, com en el cas de les variables resposta anteriors.

## CAPÍTOL 9: CONCLUSIONS

Modelar la gravetat dels accidents industrials amb substàncies perilloses ha estat molt interessant. Algunes de les conclusions a les quals s'arriba ja eren d'esperar, però aquest projecte té la avantatge de que ara l'evidència queda recolzada per aquesta anàlisi estadística.

En aquest projecte s'ha limitat l'àmbit d'estudi als accidents produïts durant el desenvolupament de tasques en instal·lacions fixes, i com a variables resposta per mesurar la gravetat dels accidents s'han utilitzat **KILLED**, que és el nombre de morts en l'accident, **INJURED**, el nombre de persones ferides en l'accident i **DAMAGE**, l'estimació econòmica del dany material.

Les variables explicatives importants per la construcció dels models han estat l'estat físic de les substàncies involucrades en l'accident, **ESTAT\_FÍSIC**, les variables que indiquen el risc de la substància o substàncies com són **HZTO**, **HZFI**, **HZEX**, **HZCO**, **HZRA**, **HZCD**, **HZAS** i **HZOX**, el tipus d'activitat, **TIPUSACT**, la causa general de l'accident, **CAUSAG**, el tipus d'accident **TIPUSACC**, el nivell de desenvolupament econòmic **MÓN** i el nombre de substàncies que han participat en l'accident, **NSUS**.

### 1. ETAPA DE DEPURACIÓ

S'han trobat molts problemes amb la base de dades, que s'han hagut de resoldre per aconseguir adaptar-la a les necessitats del projecte. Alguns dels problemes més seriosos han sigut:

- Les variables **KILLED**, **INJURED**, **EVACUATED**, **DAMAGE**, **IG** i **QY** contenien algunes estimacions a la baixa o a l'alça i no valors concrets. Això ha obligat a traduir-ho a valors estimats.
- Moltes variables, com les mencionades en el punt anterior, tenien molts zeros. Això en principi va fer sospitar que potser hi havia alguna errada i es va haver de comprovar amb la base de dades de substàncies. Aleshores es va descobrir que hi havien zeros que eren zeros però altres que eren missings. Per solucionar-ho es van fer consultes d'actualització amb el programa ACCESS.
- A més a més, també hi ha variables que contenen un gran nombre de missings, i això ha dificultat el modelat i ha fet que variables potencialment importants com **QSUS**, **PD**, **IG** i **IS** no hagin pogut formar part dels models.
- Les variables resposta contenen molts zeros i això ha dificultat el seu modelat.
- Les variables explicatives qualitatives contenien moltes categories i això va obligar a fer una recategorització. Aquest ha sigut el cas de les variables **MT**, **IT** i **OG**.
- Hi havia diferents registres amb el mateix codi d'identificació (**AN**). Alguns corresponien al mateix accident i en altres casos a accidents diferents. S'han hagut de agrupar els registres que corresponien al mateix accident i identificar amb un codi nou els accidents diferents.
- No totes les variables tenien la mateixa codificació pel que fa als missings, i això ha comportat confusions. S'ha hagut d'estandarditzar la codificació dels missings en totes les variables amb el símbol “\*”.
- Per poder comparar accidents d'anys diferents, s'han creat variables per actualitzar l'estimació del dany material a partir de l'índex d'actualització del cost Marshall & Stevens.
- La variable **COUNTRY** no indicava en alguns casos ben bé l'estat on es va produir l'accident. Per tal de tenir una variable geogràfica amb poques categories, es va crear la variable **MÓN**.

## **2. ETAPA ANÀLISI EXPLORATÒRIA DE DADES**

Pel que fa al nombre de morts, s'observa que els accidents que tenen un major nombre de morts són els produïts a la dècada dels anys 20, i en general a abans de la dècada dels 70, en els països subdesenvolupats. En quant a les substàncies que intervenen estan en estat físic pols o sòlid, quan són explosives i refrigerades, i quan hi participen una o quatre substàncies el nombre de morts també augmenta. El tipus d'accident amb més morts és l'explosió i el núvol de gas. En quant al tipus d'activitat que causa més morts és l'originat en plantes de procés o en plantes d'emmagatzematge. I per últim les causes amb més morts són per una fallida mecànica i una errada humana.

En canvi, els accidents que tenen un nombre menor de morts són els que s'han produït a la dècada dels 50, en general després dels anys 70 i en estats desenvolupats. En quant a les substàncies que intervenen en estat físic gas o gas pressuritzat, corrosives, i quan hi participen dues substàncies en l'accident el nombre de morts també disminueix. El tipus d'accident que provoca menys morts és la fuga i l'incendi, i el tipus d'activitat és l'originat durant tasques auxiliars de la indústria. Per últim, les causes amb menys morts són per un impacte i per una reacció violenta.

Pel que fa al nombre de ferits, s'observa que els accidents que tenen un major nombre de ferits són els produïts a la dècada dels anys 40, en general abans de la dècada dels 70, i en els països en via de desenvolupament. El nombre de ferits també augmenta quan les substàncies que intervenen en l'accident es troben en estat físic sòlid o líquid, són tòxiques i explosives, i quan hi participen quatre substàncies. El tipus d'accident amb més ferits és el núvol de gas i la fuga. Pel que fa al tipus d'activitat que causa més ferits és l'originat durant operacions de càrrega i descàrrega o durant tasques auxiliars de la indústria. I per finalitzar, les causes amb més ferits són per una reacció violenta i una errada humana.

D'altra banda, els accidents que tenen un nombre menor de ferits són els que s'han produït a la dècada dels 70, en general després dels anys 70, i en estats desenvolupats. Si les substàncies que intervenen es troben en estat físic gas, són asfixiants, i quan hi participen dues substàncies el nombre de ferits en l'accident també disminueix. El tipus d'accident amb menys ferits és l'incendi. El nombre de ferits també disminueix quan el tipus d'activitat és l'originat en locals domèstics o comercials i en plantes d'emmagatzematge. I per acabar, la causa que provoca menys ferits és una variació en les condicions de procés.

Pel que fa a l'estimació econòmica del dany material actualitzada amb l'índex Marshall & Stevens, s'observa que els accidents que tenen una major estimació econòmica són els produïts a la dècada dels anys 80, en general després de la dècada dels 70, i en els països subdesenvolupats. També hi ha un cost econòmic elevat quan les substàncies que intervenen es troben en estat físic gas o pols, són refrigerades, i quan hi participen quatre substàncies o una en l'accident. El tipus d'accident amb més danys és la fuga i el tipus d'activitat és l'originat en plantes de procés i en plantes d'emmagatzematge. I per últim, les causes que produeixen més danys materials són per una errada humana i per una fallida d'instruments.

En canvi, els accidents que tenen una estimació econòmica menor són els que s'han produït a la dècada dels 50, en general abans dels anys 70, i en estats en via de desenvolupament. En quant a les substàncies que intervenen es troben en estat físic sòlid o líquid, són oxidants, i quan hi participen dues substàncies en l'accident el dany material també és menor. El tipus d'accident amb menys danys és el núvol de gas i l'incendi i el tipus d'activitat és l'originat durant tasques auxiliars de la indústria. I per finalitzar, la causa que provoca un cost menor és una fallida de serveis.

### **3. ETAPA DE MODELAT**

El model lineal de regressió suposa normalitat per a la variable resposta. En aquest estudi, dues de les tres variables resposta no són contínues, sinó que són discretes i per tant els models lineals només seran vàlids com a aproximació. Aquestes dues variables són **KILLED** i **INJURED**, que són comptetjos i per tant s'haurien de modelar mitjançant models lineals generalitzats, on la variable resposta segueix una distribució de Poisson. Seguint les recomanacions fetes en aquest projecte, està previst que en un futur proper es continuï en aquesta direcció.

Per sort, sovint els models lineals es poden adaptar per aquest tipus de dades a base de transformar la variable resposta prenent logaritmes neperians, ja que això estabilitza la variància i sovint linearitza el model. Això és el que s'ha fet per solucionar-ho. Per la variable **DAMAGE**, també s'ha fet aquesta transformació perquè tot i que és una variable contínua (presenta valors propers a 1 que influeixen molt sobre la resta) té una distribució molt asimètrica i amb la transformació s'aconsegueix una distribució normal.

S'ha descartat com a variable resposta **EVACUATED**, ja que la transformació logarítmica no solucionava la no normalitat i la influència dels zeros i uns, i a més a més aquesta variable presenta molts missings i té molt soroll.

Per fer el modelat de les variables resposta, com que gairebé totes les variables explicatives són categòriques, s'ha establert un algorisme per decidir l'ordre en què es treuen del model les variables no significatives. Això condiciona el model final al qual s'arriba.

Com que les variables **KILLED** i **INJURED** tenen molts zeros, es diferencien dues poblacions. La dels accidents amb morts i la de tots els accidents (inclosos els que no tenen morts). Pel nombre de ferits es el mateix. El que s'ha fet ha estat modelar per separat els accidents amb morts i tots els accidents. Per ferits també s'ha fet el mateix.

### 3.1. MODELS PER KILLED

Els models trobats per **KILLED**, quan s'agafen tots els accidents són molts diferents als trobats pel accidents que tenen morts. Si es comparen, els que expliquen més bé la resposta i tenen una variància residual menor són els models per tots els accidents. Els models per tots els accidents quan s'inclou la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln Killed = & - 1,27 - 0,269 \text{ Eflíquid} - 0,287 \text{ HZTO} + 1,18 \text{ HZEX} + \\ & + 0,195 \text{ NSUS} + 0,974 \text{ TACCEX} + 0,975 \text{ TACTDO/C} + \\ & + 0,793 \text{ TACTPRO} + 0,508 \text{ TACTSTO} + 0,791 \text{ TACTTRANS} + \\ & + 0,531 \text{ CGHUM} - 1,08 \text{ Mdesenv} + 0,673 \text{ Menviadesenv} \end{aligned}$$

Quan no s'inclou la variable **CAUSA GENERAL** el model és:

$$\begin{aligned} \ln Killed = & - 0,880 - 0,353 \text{ Eflíquid} - 0,184 \text{ HZTO} + 1,09 \text{ HZEX} + \\ & + 1,63 \text{ HZAS} + 0,185 \text{ NSUS} + 0,927 \text{ TACCEX} - 0,352 \text{ TACCGAC} + \\ & + 0,971 \text{ TACTDO/C} + 0,691 \text{ TACTPRO} + 0,490 \text{ TACTSTO} + \\ & + 0,829 \text{ TACTTRANS} - 1,12 \text{ Mdesenv} + 0,457 \text{ Menviadesenv} \end{aligned}$$

S'observa que els dos models anteriors són diferents. Si no s'inclou la variable **CAUSA GENERAL** formen part del model variables diferents i el model que explica més, té una  $R_{adj} = 33,5\%$ , i té menys variància residual que el millor model amb la variable **CAUSA GENERAL**. Això és degut a que aquesta variable té moltes dades mancants i quan no s'inclou en el model es disposa de més observacions per explicar el nombre de morts.

L'anàlisi exploratòria ja dona una idea de les variables que expliquen el nombre de morts. A partir del model que ajusta tots els accidents i no té en compte la variable **CAUSA GENERAL** es pot concloure quines variables influeixen en el nombre de morts i en quin sentit. Cal destacar que hi ha variables explicatives correlacionades i per tant la interpretació de l'efecte de les mateixes s'ha de fer de forma conjunta.



El nombre de morts augmenta quan la substància o substàncies es troben en estat físic líquid i l'accident es produeix en una planta d'emmagatzematge. El nombre de morts també augmenta si l'accident es produeix per una explosió, en locals domèstics o comercials, o en plantes de procés, i quan hi participen substàncies explosives i/o asfixiants. A més a més, el nombre de morts augmenta en augmentar el nombre de substàncies i si l'accident es produeix durant operacions de càrrega o descàrrega. En canvi el nombre de morts disminueix quan en l'accident es produeix per un núvol de gas i hi participen substàncies tòxiques. Si l'accident es produeix en països desenvolupats el nombre de morts disminuirà més. Per contra el nombre de morts augmenta si l'accident es produeix en estats en via de desenvolupament.

### 3.2. MODELS PER INJURED

Per **INJURED** passa el mateix que amb la variable **KILLED**. Els models quan s'agafen tots els accidents són molts diferents als models trobats pels accidents que tenen ferits. Els models que expliquen més i tenen una variància residual menor són els models ajustats a tots els accidents. Si s'inclou la variable **CAUSA GENERAL** aquest és:

$$\begin{aligned} \ln \text{Injured} = & 0,931 - 0,425 \text{ Eflíquid} + 0,540 \text{ HZTO} - 0,813 \text{ HZFI} + \\ & + 0,266 \text{ NSUS} + 1,15 \text{ TACCEX} + 0,521 \text{ TACCGAC} + \\ & + 0,518 \text{ TACTDO/C} + 0,413 \text{ TACTPRO} + 0,550 \text{ TACTTRANS} + \\ & + 0,664 \text{ CGHUM} - 1,26 \text{ Mdesenv} \end{aligned}$$

Si no s'inclou la variable **CAUSA GENERAL** el model és:

$$\begin{aligned} \ln \text{Injured} = & 1,06 - 0,369 \text{ Eflíquid} + 0,555 \text{ HZTO} - 0,799 \text{ HZFI} - \\ & - 0,454 \text{ HZCO} + 0,259 \text{ NSUS} + 1,00 \text{ TACCEX} + \\ & + 0,547 \text{ TACCGAC} + 0,511 \text{ TACTDO/C} + 0,320 \text{ TACTPRO} + \\ & + 0,500 \text{ TACTTRANS} - 1,03 \text{ Mdesenv} + 0,476 \text{ Menviadesenv} \end{aligned}$$

Els dos models anteriors són diferents. Encara que els dos models expliquen el mateix,  $R_{adj} = 16,2$ , si no es té en compte la variable **CAUSA GENERAL** la variància residual és menor.

A partir del model que ajusta tots els accidents i si no es té en compte la variable **CAUSA GENERAL** es conclou que el nombre de ferits augmenta si l'accident es produeix per una explosió, en plantes de procés, o durant operacions de càrrega o descàrrega, però si intervenen substàncies inflamables l'augment serà menor. Si la substància o substàncies que intervenen en l'accident es troben en estat físic líquid, són inflamables i/o corrosives el nombre de ferits disminueix. El nombre de ferits augmenta en augmentar el nombre de substàncies. També augmenta el nombre de ferits si intervenen substàncies tòxiques i l'accident es produeix per un núvol de gas. Quan l'accident es produeix en locals domèstics o comercials el nombre de ferits augmenta. El mateix passa si l'accident es produeix en estats en via de desenvolupament. En canvi si l'accident es produeix en estats desenvolupats el nombre de ferits disminuirà més.

### 3.3. DAMAGE

Per **DAMAGE** quan s'inclou la variable **CAUSA GENERAL** el millor model és:

$$\begin{aligned} \ln \text{Damage} = & 0,925 + 0,549 \text{ Efgas} - 0,856 \text{ HZOX} + \\ & + 0,615 \text{ TACCEX} + 0,529 \text{ TACCREL} - \\ & - 1,29 \text{ TACTDO/C} + 0,387 \text{ CGEXT} \end{aligned}$$

Si no s'inclou la variable **CAUSA GENERAL** és:

$$\begin{aligned} \ln \text{Damage} = & 0,926 + 0,624 \text{ Efgas} + 0,127 \text{ NSUS} + 0,308 \text{ TACCEX} + \\ & + 1,82 \text{ TACCGAC} + 0,424 \text{ TACCREL} - 0,976 \text{ TACTDO/C} - \\ & - 1,59 \text{ TACTTRANS} - 1,57 \text{ ns*tacgc} + 1,02 \text{ ns*tatran} \end{aligned}$$

Els dos models no són iguals. Quan no es té en compte la variable **CAUSA GENERAL**, el model que explica més, té una  $R_{adj} = 7,6\%$ , i té menys variància residual. Això passa perquè aquesta variable té moltes dades mancants i quan no s'inclou en el model es disposa de més observacions per explicar l'estimació econòmica del dany material.

Donat el model sense la variable **CAUSA GENERAL**, si es té en compte que hi ha variables correlacionades es conclou que l'estimació econòmica del dany material serà menor si l'accident és una explosió i es produeix en locals domèstics o comercials, o durant operacions de càrrega o descàrrega. En canvi, si en un accident produït per una explosió hi participen substàncies en estat físic gas el cost augmentarà més. En el cas d'una substància l'estimació econòmica disminueix més si l'accident es produeix durant operacions de càrrega o descàrrega. Però en el cas d'una substància el cost augmenta si l'accident es produeix per un núvol de gas. Per acabar, el dany material augmenta si l'accident es produeix per una fuga.

### 3.4. COMPARACIÓ DELS MODELS

Comparant els models trobats per **KILLED**, **INJURED** i **DAMAGE** s'arriba a la conclusió que quan la variable **CAUSA GENERAL** (que és una variable amb moltes dades mancants) no es té en compte, poden formar part del model altres variables.

Les variables comuns que tenen tant els models per **KILLED** com els d'**INJURED** (sense incloure la variable **CAUSA GENERAL**) són: **EFLÍQUID**, **HZTO**, **NSUS**, **TACCEX**, **TACCGAC**, **TACTDO/C**, **TACTPRO**, **TACTTRANS**, **MDESENV** i **MENVIADSENV**. En els dos models totes tenen el mateix signe menys **HZTO** i **TACCGAC** que per **KILLED** són negatives i per **INJURED** positives. Això pot ser degut a la colinealitat, ja que **HZTO** i **TACCGAC** estan correlacionades. Cal destacar que el fet que hi hagi colinealitat complica la interpretació dels models.

Quan l'estat físic de la substància o substàncies que intervenen en l'accident és líquid, el nombre de morts i ferits disminueix i el mateix passa quan els accidents es produeixen en estats desenvolupats. En canvi el nombre de morts i ferits **augmenta** en augmentar el nombre de substàncies que intervenen en l'accident, quan es produeix per una explosió i/o núvol de gas, en locals domèstics o comercials o en plantes de procés o durant operacions de càrrega i descàrrega, i també quan es produeixen en estats en via de desenvolupament.

Les variables comuns que tenen els models per **KILLED**, **INJURED** i **DAMAGE** (sense la variable **CAUSA GENERAL**) són: **NSUS**, **TACCEX**, **TACCGAC**, **TACTDO/C** i **TACTTRANS**. Les variables que tenen el mateix signe són **NSUS** i **TACCEX**, és a dir, el nombre de morts, el de ferits i l'estimació econòmica del dany material augmenta conforme augmenta el nombre de substàncies que intervenen en l'accident i quan aquest es produeix per una explosió. Les altres variables tenen signe diferent segurament degut a la colinealitat que hi ha entre aquestes variables o perquè una variable que pot causar ferits o morts no té perquè causar tants danys materials.

El fet que les variables regressores expliquin poc és degut al nivell de soroll de les variables resposta, a la imprecisió de la base de dades i també pot ser degut a que hi ha variables importants (com per exemple, distribució en planta, orografia del terreny i equips d'intervenció) que expliquen aquestes variables resposta i que no estan a l'abast. La utilització dels models per tant té més sentit quan l'objectiu és explicar relacions entre respostes i explicatives, que quan és el de predir.

# VALORACIÓ TEMPORAL I ECONÒMICA

El cost econòmic d'aquest projecte està calculat per persona. El projecte es va començar a mitjans del mes de febrer i s'ha acabat a finals del mes de juny. En total s'han dedicat 380 hores per persona i han estat repartides de la següent manera:

**1) Depuració de la base de dades:**

$$70 \text{ hores} \times 2.500 \text{ ptes/hora} = 175.000 \text{ ptes.}$$

**2) Reunions amb l'enginyer responsable del projecte:**

$$30 \text{ hores} \times 1.500 \text{ ptes/hora} = 45.000 \text{ ptes.}$$

**3) Anàlisi de dades:**

$$200 \text{ hores} \times 2.500 \text{ ptes/hora} = 500.000 \text{ ptes.}$$

**4) Redacció de l'informe:**

$$80 \text{ hores} \times 2.000 \text{ ptes/hora} = 160.000 \text{ ptes.}$$

Per tant, el cost total per persona és:

$$175.000 + 45.000 + 500.000 + 160.000 = \underline{\underline{880.000 \text{ ptes.}}}$$

El cost total de l'estudi és:

$$880.000 \text{ ptes/persona} \cdot 2 \text{ persones} = \underline{\underline{1.760.000 \text{ ptes.}}}$$

## **ANNEX A: MOSTRA DE LA BASE DE DADES MHIDAS**

En aquest annex es recull una mostra de la base de dades de substàncies MHIDAS de 17 observacions. Per tant, només és un subconjunt de les 10.102 substàncies registrades a la base de dades.

| AN     | CR  | DIA | MES | AÑO | PD | LOCATION       | REGION          | COUNTRY     | MN             | MT         | MH     | MC   | IT                |     |
|--------|-----|-----|-----|-----|----|----------------|-----------------|-------------|----------------|------------|--------|------|-------------------|-----|
| 1004 A | SRD | 05  | 10  | 84  | T  | PORT TALBOT    | WEST GLAMORGAN  | UK          | PROPANE        | PLGAS      | FI-    | 1075 | FIRE-EXPLODE-     | TR/ |
| 1006   | SRD | 18  | 08  | 84  | T  | JACKSONVILLE   | MASSACHUSETTS   | USA         | GASOLINE       | LIQUID     | FI-    | 1270 | FIRE-FIRE-        | ST  |
| 1011   | SRD | 15  | 06  | 84  | R  | MILFORD HAVEN  | DYFED           | UK          | DIESEL OIL     | LIQUID     | FI-    | 1223 | EXPLODE-          | T   |
| 1013 A | SRD |     |     | 84  |    |                | CLEVELAND       | UK          | CL FL PICOLENE | PLGAS      | TO-    | 2313 | EXPLODE-BLEVE-    | WA  |
| 1013 B | SRD |     |     | 84  |    |                | CLEVELAND       | UK          | BITUMEN        | LIQUID     | FI-    | 1999 | FIRE-             | V   |
| 1013 C | SRD |     |     | 84  |    |                | CLEVELAND       | UK          | LPG            | PLGAS      | FI-    | 1075 | EXPLODE-BLEVE-    | V   |
| 1014   | SRD | 13  | 10  | 84  |    | BROMBOROUGH    | MERSEYSIDE      | UK          | OIL/DUST/FIBRE | DUST       | FI-    | 9999 | FIRE-             |     |
| 1015   | SRD | 28  | 10  | 73  |    | NAKAKUBIKI     | NIIGATA         | JAPAN       | VINYL CHLORIDE | LIQUID     | TO-FI- | 1086 | RELEASE-FIRE-     | TR/ |
| 1017   | SRD | 09  | 06  | 84  | R  | NORTH BAY      | ONTARIO         | CANADA      | SULPHURIC ACID | LIQUID     | CO-OX- | 1830 | ELEASE-CONTREIRAN |     |
| 1018   | SRD | 25  | 10  | 84  | R  | SYDNEY         | NEW SOUTH WALES | AUSTRALIA   | BUNKER OIL     | LIQUID     | FI-    | 1267 | ELEASE-CONTREL-TR |     |
| 1019   | SRD | 08  | 06  | 84  |    | MARL           |                 | GERMANY     |                |            |        | 9999 | ELEASE-CONTREL-PR |     |
| 1273 A | SRD | 28  | 12  | 69  |    | FAWLEY         | HAMPSHIRE       | UK          | HYDROGEN       | GAS        | FI-    | 1049 | RELEASE-FIRE-PR   |     |
| 1273 B | SRD | 28  | 12  | 69  |    | FAWLEY         | HAMPSHIRE       | UK          | NAPHTHA        | SUPERHEATL | FI-    | 1270 | RELEASE-FIRE-PR   |     |
| 1274   | SRD | 06  | 03  | 69  |    | PUERTO LA CRUZ |                 | VENEZUELA   | HYDROCARBONS   | LIQUID     | FI-    | 1223 | EXPLODE-          |     |
| 1275   | SRD | 14  | 05  | 69  |    | WILTON         | CLEVELAND       | UK          | CYCLOHEXANE    | LIQUID     | FI-    | 1145 | GASCLD-FIRE-PR    |     |
| 1277   | SRD | 16  | 08  | 69  |    | SILVERWATER    | SYDNEY          | AUSTRALIA   | CHEMICALS      |            | FI-    | 9999 | EXPLODE-FIRE-     |     |
| 1279   | SRD | 10  | 11  | 69  |    | BOTLEK         |                 | NETHERLANDS | NAPHTHA        | LIQUID     | FI-    | 1270 | FIRE-             |     |

| AN     | OG                   | GC                         | SC         | QY    | IS             | IG  | DAMAGE        | KILLED | INJURED | EVACUATED     |
|--------|----------------------|----------------------------|------------|-------|----------------|-----|---------------|--------|---------|---------------|
| 1004 A | TRANSPORT;PACKAGE    | EXTERNAL-                  | EXTNLFIRE- |       | FLAMES-BOILER  |     |               | 0      | 0       | 0 x 10E3      |
| 1006   | STORAGE; ASVESSEL    | EXTERNAL-                  | LIGHTNING- | >3000 | ELECTRIC-STATI |     |               | 0      | 0       | 0 x 10E3      |
| 1011   | TRANSPORT; SHIP      | HUMAN-PROCONIERALOP-DRAIN  |            | 6     | FLAMES-FLARE   |     |               | 4      | 17      |               |
| 1013 A | WASTE;ROADTANKER     | EXTERNAL-                  | EXTNLFIRE- | 1     |                |     |               |        |         |               |
| 1013 B | WASTE; PACKAGE       | EXTERNAL-                  | EXTNLFIRE- | 1     |                |     |               | 0      | 0       |               |
| 1013 C | WASTE;PSVESSEL       | EXTERNAL-                  | EXTNLFIRE- | 1     |                |     |               | 0      | 0       |               |
| 1014   | PROCESS;PUMP         | VREACTION-                 | INTNLFIRE- |       |                |     | 6.7 x 10E6    |        | 2       | >0.001 x 10E3 |
| 1015   | TRANSFER;PIPEWORK    | HUMAN-MECHANIC;NERALOP-VAL |            | 4.2   |                | 900 |               | 1      | 23      |               |
| 1017   | TRANSPORT;RAILTANKER | IMPACT-                    | RAILACC-   | 993   |                |     |               |        |         |               |
| 1018   | TRANSPORT;PIPELINE   | IMPACT-                    | VEHICLE-   |       |                |     |               | 0      |         |               |
| 1019   | PROCESS;PIPEWORK     |                            |            |       |                |     | 4.6 x 10E6    |        |         |               |
| 1273 A | PROCESS;PIPEWORK     | MECHANICAL-                |            | 23    |                |     | 3.5 x 10E6    | 0      | 4       |               |
| 1273 B | PROCESS;PIPEWORK     | MECHANICAL-                |            | 23    |                |     | 3.5 x 10E6    | 0      | 4       |               |
| 1274   | PROCESS              |                            |            |       |                |     | >0.5 x 10E6   | 5      | >15     |               |
| 1275   | PROCESS;PIPEWORK     | HUMAN-                     | INSTALL-   | 4     | ELECTRIC-VEHIC |     | 18 x 10E6     | 2      | 23      |               |
| 1277   | PROCESS;PUMP         |                            |            |       |                |     |               | <3     | <3      |               |
| 1279   | PROCESS;PUMP         | MECHANICAL-                | GLANDSEAL- |       |                |     | <0.055 x 10E6 | 0      | 0       |               |



| AN     | KW   | AB   | RA |
|--------|--|--|----|
| 1004 A | DURATION-Duration of the incident; RESPONSE-Emergency response to the incident   | IRON + STEELWORKS. 4 MEN ON ROOF RENOVATING. MOLTEN BITUMEN BOILED OVER OUT OF BOILERS IGNITED BY PROPANE BURNERS. WORKMAN ATTEMPTED TO EXTINGUISH BUT FAILED. PROPANE CYLINDERS EXPLODED, INTENSIFIED BLAZE. FIRE SERVICE CONTROLLED FIRE IN 30 MINS.     | 1  |
| 1006   | COMPANY-Firm or company involved; DISPERSE-Dispersion of material; DURATION-Duration of the incident; ENVIRMNT-Environmental impact of the incident (natural environment); RESPONSE-Emergency response to the incident   | LIGHTNING STRIKE ON OIL TANK FARM CAUSED FIRE THAT EVENTUALLY MELTED THE TANK CAUSING IT TO RUPTURE. THE REMAINING GASOLINE SPILLED OUT AND A LARGE FIREBALL FORMED. COMPANY WAS TRIANGLE OIL CORP.  | 1  |
| 1011   | ADVICE-Contains recommendations to prevent repetitions of the incident; DISPERSE-Dispersion of material; ENQUIRY-Investigations, public enquiries, etc., into the incident; PRODESC-Contains a description of the process as usually undertaken at the site of | TANKER "POINTSMAN" IN MILFORD HAVEN FOR MAINTENANCE CERTIFIED GAS FREE. EXPLOSION OCCURRED WHEN WORKERS LIT ACETYLENE TORCH IGNITING GAS OIL/MOTOR SPIRIT VAPOURS. 2 FURTHER EXPLOSIONS INJURED FIREFIGHTERS.  | 4  |
| 1013 A | PROGRESN-How the incident progressed or developed with time  | GRASS FIRE SPREAD TO CHEMICAL WASTE DISPOSAL DEPOT IGNITING TIMBER PALLETS AND BITUMEN DRUMS (45gall) BLEVE OF LPG STORAGE TANK DID MUCH TO EXTINGUISH FIRE. ONE OF NEARBY TANKERS BORE LABEL FOR 7017 "CHLOROFLUOROPI RESIDUES"                           | 1  |
| 1013 B | PROGRESN-How the incident progressed or developed with time  | GRASS FIRE SPREAD TO CHEMICAL WASTE DISPOSAL DEPOT IGNITING TIMBER PALLETS AND BITUMEN DRUMS (45gall) BLEVE OF LPG STORAGE TANK DID MUCH TO EXTINGUISH FIRE.   | 1  |
| 1013 C | PROGRESN-How the incident progressed or developed with time  | GRASS FIRE SPREAD TO CHEMICAL WASTE DISPOSAL DEPOT IGNITING TIMBER PALLETS AND BITUMEN DRUMS (45gall) BLEVE OF LPG STORAGE TANK DID MUCH TO EXTINGUISH FIRE.   | 1  |
| 1014   | DESIGN-Design of plant equipment; DURATION-Duration of the incident; PRODESC-Contains a description of the process as usually undertaken at the site of the incident; RESPONSE-Emergency response to the incident  | A FIRE BEGAN IN A FILTER UNIT AND THIS SPREAD TO THE ROOF BEFORE APPARENTLY BEING EXTINGUISHED BY PRODUCTION OPERATORS. 25 MINS LATER A FIRE WAS DISCOVERED IN STORAGE AREA, PREMISES EVACUATED AND FIRE BRIGADE CALLED.                                   | 1  |
| 1015   | COMPANY-Firm or company involved; TIMING-Timing of the incident, time of day, shift, handover, lunch break, etc.; DURATION-Duration of the incident; ENQUIRY-Investigations, public enquiries, etc., into the incident   | VINYL CHLORIDE MONOMER LEAKED & CAUGHT FIRE WHEN STRAINER VALVE ON MONOMER LINE FAILED. REPORT BY ACCIDENT SURVEY TEAM INDICATED THAT OPERATOR DAMAGED VALVE WHEN TIGHTENING WITH AN UNAUTHORISED TOOL. FIRE LASTED 45hrs. COMPANY WAS SHINETSU CHEMICALS. | 1  |
| 1017   | AFTER-After effects such as clear up, survival of injured, etc.; ENVIRMNT-Environmental impact of the incident (natural environment); TERRAIN-Physical geography in the area of the incident   | DERAILMENT OF 57 TANKER CARS LOADED WITH 93% SULPHURIC ACID LED TO FISH DYING AS 390,595 US GALLS SPILLED INTO RIVER.  | 1  |
| 1018   | ENVIRMNT-Environmental impact of the incident (natural environment); FACTORS-Aggravating or mitigating factors; NUISANCE-Nuisance (non-hazardous) caused by the incident; PROGRESN-How the incident progressed or developed with time                          | DREDGER SHIP WORKING NEAR CALTEX REFINERY WHARF FRACTURED SUBMERGED BUNKER PIPELINE GIVING MINOR SPILLAGE. DURING CLEAN-UP, FLUSHING PIPELINE CAUSED CONSIDERABLE POLLUTION OF BEACHES IN AREA.  | 1  |
| 1019   | COMPANY-Firm or company involved   | FIRE AFTER PIPE BROKE AND PRODUCT ESCAPED FROM CHEMISCHE WERKE HULS VEBA OIL.  | 1  |
| 1273 A | COMPANY-Firm or company involved; MAP-Hard copy contains map; RESPONSE-Emergency response to the incident  | FAILURE OF 400mm PIPE AT 1.7mpa RELEASED LARGE QUANTITIES OF MATERIAL FROM HYDROFINING SECTION OF OIL REFINERY. IGNITION AT UNDETERMINED SOURCE LED TO INTENSE FIRE & FURTHER RELEASE OF NAPHTHA WHICH FUELLED BLAZE.                                      | 2  |
| 1273 B | COMPANY-Firm or company involved; MAP-Hard copy contains map; RESPONSE-Emergency response to the incident  | FAILURE OF 400mm PIPE AT 1.7mpa RELEASED LARGE QUANTITIES OF MATERIAL FROM HYDROFINING SECTION OF OIL REFINERY. IGNITION AT UNDETERMINED SOURCE LED TO INTENSE FIRE & FURTHER RELEASE OF NAPHTHA WHICH FUELLED BLAZE.                                      | 2  |
| 1274   | COMPANY-Firm or company involved; TIMING-Timing of the incident, time of day, shift, handover, lunch break, etc.   | AN EXPLOSION DESTROYED A LARGE PART OF A GULF PETROLEUM REFINERY. PRODUCTION HALTED FOR UNKNOWN PERIOD. TERRORISM RULED OUT. EXTENSIVE DAMAGE TO HOUSES, WAREHOUSES + OFFICES NEARBY. ESTIMATES OF DAMAGE INITIALLY >0.5 MILLION DOLLARS.                  | 4  |
| 1275   | COMPANY-Firm or company involved; ADVICE-Contains recommendations to prevent repetitions of the incident   | A LOW CLOUD FORMED AT 689KPa FROM AN IMPROPERLY BLANKED PIPE. GAS CLOUD IGNITED BY DIESEL ENGINE 30m FROM SPILL POINT. ATTEMPTS TO STOP ENGINE FAILED SINCE ENGINE RUNNING ON CYCLOHEXANE SUCKED IN THROUGH AIR INTAKE.                                    | 3  |
| 1277   | COMPANY-Firm or company involved   | EYEWITNESSES SAID EXPLOSION OCCURRED IN A COMPRESSOR UNIT. FIRE UNDER CONTROL 45 MINS. LATER. FLAMMABLE GASES UNDER HIGH PRESSURE IGNITED AND FLAMES WENT 90'HIGH. 3 MEN MISSING- PRESUMED DEAD BUT UNCERTAIN FROM REPORTS. CO = PACCAL CORP.              | 1  |
| 1279   | COMPANY-Firm or company involved; TIMING-Timing of the incident, time of day, shift, handover, lunch break, etc.; DURATION-Duration of the incident  | FIRE IN NAPHTHA CRACKING UNIT THOUGHT TO BE DUE TO LEAK FROM PUMP. FIRE EXTINGUISHED AFTER 3hrs.   | 1  |

## **ANNEX B: MOSTRA DE LA BASE DE DADES MHIDAS ADAPTADA**

En aquest annex es recull una mostra de la base de dades d'accidents adaptada a les necessitats de l'estudi de 44 observacions. Només conté un subconjunt dels 5.167 accidents amb substàncies perilloses que formen la base de dades adaptada.

| AN | CR  | DIA | MES | ANY | PD | LOCATION       | REGION         | COUNTRY      | MN                                  | ESTAT_FÍSIC  | HZTO | HZFI | HZEX | HZCO |
|----|-----|-----|-----|-----|----|----------------|----------------|--------------|-------------------------------------|--------------|------|------|------|------|
| 1  | SRD | 28  | 4   | 87  | T  | WAKEFIELD      | YORKSHIRE      | UK           | CYANURIC CHLORIDE/CYANURIC CHLORIDE | GAS/GAS      | 0    | 0    | 0    | 1    |
| 3  | SRD | 27  | 1   | 81  |    | ANTWERP        | ANTWERPEN      | BELGIUM      | MALT                                | DUST         | 0    | 0    | 1    | 0    |
| 6  | SRD | 30  | 11  | 72  | T  | BARCARENA      |                | PORTUGAL     | GUNPOWDER                           | DUST         | 0    | 0    | 1    | 0    |
| 8  | SRD | 1   | 12  | 75  |    | BUSAN          | SEOUL          | KOREA        | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 10 | SRD | 1   | 8   | 61  | T  | CROVE COCUB    | ILLINOIS       | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 12 | SRD | 23  | 2   | 74  |    | MARSHALL       | MINNESOTA      | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 13 | SRD | 4   | 5   | 71  | R  | ANTWERP        | ANTWERPEN      | BELGIUM      | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 14 | SRD | 15  | 3   | 71  | T  | SOUTHWARK      | GREATER LONDON | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 15 | SRD | 22  | 11  | 73  |    | GLENROTHES     | FIFE           | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 16 | SRD | 28  | 8   | 64  | T  | FRANKFURT      |                | GERMANY      | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 17 | SRD | 20  | 7   | 64  | T  | LEITH          | LOTHIAN        | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 19 | SRD | 23  | 4   | 70  | T  | WILTON         | CLEVELAND      | UK           | AMMONIA                             |              | 1    | 1    | 0    | 0    |
| 20 | SRD | *   | 11  | 77  |    | PALEMBANG      | SUMATRA        | INDONESIA    | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 21 | SRD | 18  | 6   | 77  | T  | LLANDARCY      | W.GLAMORGAN    | UK           | AMMONIA                             | GAS          | 1    | 1    | 0    | 0    |
| 23 | SRD | 4   | 9   | 64  | T  | WILTON         | CLEVELAND      | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 25 | SRD | 23  | 10  | 66  |    | BOLOEIL        | QUEBEC         | CANADA       | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 26 | SRD | *   | *   | 57  |    | URALS          |                | CIS          | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 27 | SRD | *   | 7   | 55  |    | EAST HAM       | GREATER LONDON | UK           | AMMONIA                             | LGAS         | 1    | 1    | 0    | 0    |
| 32 | SRD | 7   | 5   | 76  | T  | HEXHAM         | NORTHUMBERLAND | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 33 | SRD | 9   | 9   | 82  | T  | TORONTO        | ONTARIO        | CANADA       | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 37 | SRD | 16  | 5   | 80  | V  | CORK           | CORK           | EIRE         | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 40 | SRD | *   | *   | 77  |    | OHIO RIVER     | OHIO           | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 41 | SRD | 18  | 10  | 72  |    | SVELGEN        |                | NORWAY       |                                     | DUST         | 0    | 0    | 0    | 0    |
| 42 | SRD | 9   | 12  | 77  |    | CARTAGENA      |                | COLOMBIA     | AMMONIA                             | GAS          | 1    | 1    | 0    | 0    |
| 44 | SRD | 13  | 7   | 73  | T  | POTCHEFSTROOM  | NATAL          | SOUTH AFRICA | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 45 | SRD | 30  | 5   | 73  | R  | INCE MARSHES   | CHESHIRE       | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 47 | SRD | 14  | 8   | 72  |    | BILLINGS       | MONTANA        | USA          | BUTANE                              | LGAS         | 0    | 1    | 0    | 0    |
| 48 | SRD | 22  | 2   | 82  |    | TAFT           | LOUISIANA      | USA          | AMMONIA/SULPHURIC ACID              | PLGAS/LIQUID | 1    | 1    | 0    | 1    |
| 49 | SRD | 13  | 7   | 77  | T  | YORK           | YORKSHIRE      | UK           | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 50 | SRD | 7   | 3   | 77  | T  | CUERNAVACA     |                | MEXICO       | AMMONIA                             | GAS          | 1    | 1    | 0    | 0    |
| 51 | SRD | 16  | 5   | 80  | V  | CORK           |                | EIRE         | AMMONIA                             | GAS          | 1    | 1    | 0    | 0    |
| 52 | SRD | 11  | 4   | 59  | R  | SHELLHAVEN     | ESSEX          | UK           | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 53 | SRD | *   | *   | 61  |    | PASADENA       | TEXAS          | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 54 | SRD | *   | 1   | 64  |    |                |                | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 55 | SRD | 11  | 7   | 59  |    | UBE            |                | JAPAN        | HYDROGEN                            |              | 0    | 1    | 0    | 0    |
| 56 | SRD | 3   | 8   | 65  |    | NEW ORLEANS    | LOUISIANA      | USA          | AMMONIA                             |              | 1    | 1    | 0    | 0    |
| 57 | SRD | 17  | 12  | 84  | T  | MATAMOROS      |                | MEXICO       | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 58 | SRD | 10  | 6   | 81  |    | BARNESVILLE    | MINNESOTA      | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 59 | SRD | *   | *   | 65  |    | PASADENA       | TEXAS          | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 60 | SRD | 8   | 1   | 82  |    |                |                | ZAMBIA       | AMMONIA                             | GAS          | 1    | 1    | 0    | 0    |
| 61 | SRD | 28  | 6   | 77  |    | LASALLE        | QUEBEC         | CANADA       | CHLORINE                            |              | 1    | 0    | 0    | 0    |
| 62 | SRD | 19  | 12  | 74  |    | NORTH TEES     | TYNE AND WEAR  | UK           | AMMONIA                             | RLGAS        | 1    | 1    | 0    | 0    |
| 63 | SRD | *   | *   | 64  |    |                |                | DENMARK      | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |
| 64 | SRD | 14  | 6   | 65  |    | CHESAPEAKE BAY | MARYLAND       | USA          | AMMONIA                             | PLGAS        | 1    | 1    | 0    | 0    |

| AN | HZRA | HZCD | HZAS | HZOX | MC        | NSUS | TIPUSACC         | IT2               | TIPUSACT   | OG2         | CAUSAG                 | SC                       |
|----|------|------|------|------|-----------|------|------------------|-------------------|------------|-------------|------------------------|--------------------------|
| 1  | 0    | 0    | 0    | 0    | 2670/2670 | 2    | RELEASE-GASCLD-  | RELEASE-GASCLD-   | DOM/COM-   | PACKAGE-    | HUMAN-VREACTION-       | PROCEDURES-              |
| 3  | 0    | 0    | 0    | 0    | 9999      | 1    | EXPLODE-         | DUSTEXP-          | TRANSFER-  | ASVESSEL-   | VREACTION-             | CONEXP-                  |
| 6  | 0    | 0    | 0    | 0    | 27        | 1    | EXPLODE-         | EXPLODE-          | PROCESS-   |             |                        |                          |
| 8  | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | INSTREL-          | DOM/COM-   | PSVESSEL-   |                        |                          |
| 10 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-GASCLD-   | TRANSFER-  | BARGE-      | MECHANICAL-            | HOSE-                    |
| 12 | 0    | 0    | 0    | 0    | 1005      | 1    | FIRE-            | FIRE-             | DOM/COM-   | PIPEWORK-   |                        |                          |
| 13 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-GASCLD-   | TRANSFER-  | PACKAGE-    | MECHANICAL-            | GLANDSEAL-VALVE-         |
| 14 | 0    | 0    | 0    | 0    | 1005      | 1    | EXPLODE-GASCLD-  | EXPLODE-GASCLD-   | WAREHOUSE- | PACKAGE-    | PROCOND-EXTERNAL-      | EXTNLFIRE-EXTNLEXP-      |
| 15 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | INSTREL-GASCLD-   | DOM/COM-   | PACKAGE-    | IMPACT-                |                          |
| 16 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | RELEASE-DENSGSCL- | TRANSFER-  | SHIP-       | MECHANICAL-HUMAN-      | CONNECTING-              |
| 17 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | RELEASE-GASCLD-   | TRANSFER-  | RAILTANKER- | MECHANICAL-            |                          |
| 19 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | RELEASE-          | PROCESS-   |             | MECHANICAL-            | VALVE-                   |
| 20 | 0    | 0    | 0    | 0    | 1005      | 1    |                  |                   | STORAGE-   | ASVESSEL-   | MECHANICAL-            | OVERPRES-                |
| 21 | 0    | 0    | 0    | 0    | 1005      | 1    | FIRE-RELEASE-    | FIRE-RELEASE-     | PROCESS-   | PVESSEL-    |                        |                          |
| 23 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | RELEASE-DENSGSCL- | TRANSFER-  | HOSE-       | MECHANICAL-            |                          |
| 25 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-GASCLD-   | TRANSFER-  | RAILTANKER- | MECHANICAL-            |                          |
| 26 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-EXPLODE- | CONTREL-EXPLODE-  | PROCESS-   |             |                        |                          |
| 27 | 0    | 0    | 0    | 0    | 1005      | 1    |                  |                   | STORAGE-   | ASVESSEL-   | VREACTION-             | CONEXP-                  |
| 32 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | DOM/COM-   | PIPEWORK-   | HUMAN-                 |                          |
| 33 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-GASCLD-   | DOM/COM-   | PIPEWORK-   | MECHANICAL-            | VALVE-                   |
| 37 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-DENSGSCL- | PROCESS-   | PIPEWORK-   | MECHANICAL-            | RELIEFVALV-              |
| 40 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-GASCLD-   | TRANSFER-  | HOSE-       | MECHANICAL-            | METALLURG-HOSE-          |
| 41 | 0    | 0    | 0    | 0    | 9999      | 1    | EXPLODE-         | DUSTEXP-          | PROCESS-   | SIZECHANGE- |                        |                          |
| 42 | 0    | 0    | 0    | 0    | 1005      | 1    | EXPLODE-RELEASE- | EXPLODE-RELEASE-  | PROCESS-   | REACTOR-    | EXTERNAL-VREACTION-    | EXTNLEXP-CONEXP-         |
| 44 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | INSTREL-DENSGSCL- | TRANSFER-  | PSVESSEL-   | MECHANICAL-            | BRITTLE-WELDFAIL-        |
| 45 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | RELEASE-GASCLD-   | TRANSFER-  | PIPEWORK-   | HUMAN-                 | ISOLUNCOUP-              |
| 47 | 0    | 0    | 0    | 0    | 1075      | 1    | FIRE-EXPLODE-    | FIRE-EXPLODE-     | PROCESS-   |             | MECHANICAL-HUMAN-      | VALVE-ISOLUNCOUP-        |
| 48 | 0    | 0    | 0    | 1    | 1005/1830 | 2    | RELEASE-GASCLD-  | CONTREL-GASCLD-   | TRANSFER-  | PIPELINE-   | IMPACT-                | SHIP/LAND-               |
| 49 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | RELEASE-          | DOM/COM-   | PIPEWORK-   | MECHANICAL-            |                          |
| 50 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-DENSGSCL- | PROCESS-   | PIPEWORK-   |                        |                          |
| 51 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | PROCESS-   | PIPEWORK-   | HUMAN-MECHANICAL-      | OVERPRES-                |
| 52 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | TRANSFER-  | HOSE-       | MECHANICAL-            | HOSE-                    |
| 53 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | TRANSFER-  | HOSE-       | MECHANICAL-            | HOSE-                    |
| 54 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | TRANSFER-  | HOSE-       | MECHANICAL-            | HOSE-                    |
| 55 | 0    | 0    | 0    | 0    | 1049      | 1    | EXPLODE-         | EXPLODE-          | PROCESS-   | PVESSEL-    |                        |                          |
| 56 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-FIRE-    | CONTREL-FIRE-     | DOM/COM-   | PIPEWORK-   | IMPACT-                | HVYOBJECT-               |
| 57 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | RELEASE-GASCLD-   | TRANSFER-  | HOSE-       | MECHANICAL-            | OVERPRES-                |
| 58 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-GASCLD-  | CONTREL-DENSGSCL- | TRANSFER-  | PSVESSEL-   | INSTRUMENT-HUMAN-      | TRIP-GENERALOP-OVERFILL- |
| 59 | 0    | 0    | 0    | 0    | 1005      | 1    | EXPLODE-         | EXPLODE-          | PROCESS-   | PVESSEL-    | MECHANICAL-INSTRUMENT- | OVERPRES-CONTROL-        |
| 60 | 0    | 0    | 0    | 0    | 1005      | 1    | EXPLODE-FIRE-    | EXPLODE-FIRE-     | PROCESS-   | PVESSEL-    |                        |                          |
| 61 | 0    | 0    | 0    | 0    | 1017      | 1    | EXPLODE-FIRE-    | EXPLODE-FIRE-     | WAREHOUSE- | PACKAGE-    |                        |                          |
| 62 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | TRANSFER-  | PIPEWORK-   | MECHANICAL-            |                          |
| 63 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-         | CONTREL-          | STORAGE-   | PSVESSEL-   | MECHANICAL-            | CORRODE-                 |
| 64 | 0    | 0    | 0    | 0    | 1005      | 1    | RELEASE-RELEASE- | RELEASE-POOL-     |            |             | IMPACT-                | ROADACC-VEHICLE-         |

| AN | QSUS     | QY3 | IS                | IG | DAMAGE1 | DG3 | DAMAGE-CEI | DAMAGE | INDEX-CEI | INDEX-MAR | KILLED | NPM3 | INJURED | NPH3 | EVACUATED | NPE3 | MÓN | RA |
|----|----------|-----|-------------------|----|---------|-----|------------|--------|-----------|-----------|--------|------|---------|------|-----------|------|-----|----|
| 1  | *        | *   |                   | *  | *       | *   | *          | *      | 0         | 0         | 0      | 3    | 5       | 3    | *         | *    | 1   | 1  |
| 3  | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 1   | 1  |
| 6  | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | 7      | 3    | *       |      | *         |      | 1   | 1  |
| 8  | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 5      | 3    | 85      | 3    | *         |      | 3   | 1  |
| 10 | 350,0000 | 3   | NONIGNITE-        | *  | *       | *   | *          | *      | 0         | 0         | *      | *    | *       | *    | 10000     | 3    | 1   | 1  |
| 12 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 1   | 1  |
| 13 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | 17      | 3    | *         |      | 1   | 1  |
| 14 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 10      | 3    | *         |      | 1   | 1  |
| 15 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 24      | 3    | *         |      | 1   | 1  |
| 16 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 3      | 3    | 20      | 3    | *         |      | 1   | 1  |
| 17 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 0       | 3    | *         |      | 1   | 1  |
| 19 | 2,0000   | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | 1       | 3    | *         |      | 1   | 1  |
| 20 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 0       | 3    | *         |      | 3   | 1  |
| 21 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | *      |      | 0       | 3    | *         |      | 1   | 1  |
| 23 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 1   | 1  |
| 25 | 15,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 1   | 1  |
| 26 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 3   | 1  |
| 27 | 581,0000 | 1   |                   | *  | *       |     | *          | *      | 0         | 0         | *      |      | 4       | 3    | *         |      | 1   | 1  |
| 32 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 1      | 3    | *       |      | *         |      | 1   | 1  |
| 33 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 1      | 3    | 3       | 3    | 550       | 3    | 1   | 2  |
| 37 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 0       | 3    | *         |      | 1   | 1  |
| 40 | 25,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | 33      | 3    | *         |      | 1   | 1  |
| 41 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | 2      | 3    | 1       | 2    | *         |      | 1   | 1  |
| 42 | *        |     |                   | *  | 2       | 3   | 3,841      | 4,234  | 204,1     | 505,4     | 21     | 3    | 30      | 3    | *         |      | 3   | 4  |
| 44 | 38,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 18     | 3    | 65      | 3    | *         |      | 2   | 3  |
| 45 | 4,0000   | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 0       | 3    | *         |      | 1   | 1  |
| 47 | *        |     |                   | *  | 5       | 3   | 14,286     | 16,113 | 137,2     | 332       | *      |      | *       |      | *         |      | 1   | 1  |
| 48 | 48,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 1       | 3    | *         |      | 1   | 2  |
| 49 | 3,0000   | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 1      | 3    | 18      | 3    | *         |      | 1   | 1  |
| 50 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 2      | 3    | 500     | 3    | 2000      | 3    | 2   | 3  |
| 51 | 10,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 0       | 3    | *         |      | 1   | 1  |
| 52 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | 3       | 3    | *         |      | 1   | 1  |
| 53 | 20,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 1   | 1  |
| 54 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 2      | 3    | 2       | 3    | *         |      | 1   | 1  |
| 55 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | 11     | 3    | 40      | 3    | *         |      | 1   | 1  |
| 56 | *        |     | ELECTRIC-MOTOREN- | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 1   | 1  |
| 57 | 30,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 182     | 3    | 3000      | 3    | 2   | 1  |
| 58 | 45,0000  | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 30      | 3    | *         |      | 1   | 2  |
| 59 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | 2      | 3    | 3       | 3    | *         |      | 1   | 2  |
| 60 | *        |     |                   | *  | *       |     | *          | *      | 0         | 0         | *      |      | *       |      | *         |      | 3   | 1  |
| 61 | *        |     |                   | *  | 1       | 3   | 1,921      | 2,117  | 204,1     | 505,4     | *      |      | *       |      | *         |      | 1   | 1  |
| 62 | 0,0000   | 3   | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 1       | 3    | *         |      | 1   | 1  |
| 63 | *        |     | NONIGNITE-        | *  | *       |     | *          | *      | 0         | 0         | 0      | 3    | 0       | 3    | 0         | 3    | 1   | 1  |
| 64 | 3,0000   | 3   | NONIGNITE-        | 0  | *       |     | *          | *      | 0         | 0         | 4      | 3    | *       |      | *         |      | 1   | 1  |

## **ANNEX C: MOSTRA DE LA BASE DE DADES UTILITZADA PER LA MODELITZACIÓ**

En aquest annex es recull una mostra de la base de dades d'accidents adaptada amb les variables resposta i explicatives que s'utilitzen per construir els models lineals. Només conté un subconjunt dels 5.167 accidents amb substàncies perilloses que formen la base de dades adaptada.

| AN | InKilled | InInjured | InEvacuated | InDamage | Efsolid | Efliquid | Efgas | Efpigas | Efdust | HZTO | HZFI | HZEX | HZCO | HZRA | HZCD | HZAS | HZOX | NSUS |
|----|----------|-----------|-------------|----------|---------|----------|-------|---------|--------|------|------|------|------|------|------|------|------|------|
| 1  | -2,30259 | 1,62924   | *           | *        | 0       | 0        | 1     | 0       | 0      | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 2    |
| 3  | *        | *         | *           | *        | 0       | 0        | 0     | 0       | 1      | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 1    |
| 6  | 1,96009  | *         | *           | *        | 0       | 0        | 0     | 0       | 1      | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 1    |
| 8  | 1,62924  | 4,44383   | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 10 | *        | *         | 9,2104      | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 12 | *        | *         | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 13 | *        | 2,83908   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 14 | -2,30259 | 2,31254   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 15 | -2,30259 | 3,18221   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 16 | 1,1314   | 3,00072   | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 17 | -2,30259 | -2,30259  | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 19 | *        | 0,09531   | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 20 | -2,30259 | -2,30259  | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 21 | *        | -2,30259  | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 23 | *        | *         | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 25 | *        | *         | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 26 | *        | *         | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 27 | *        | 1,41099   | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 32 | 0,09531  | *         | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 33 | 0,09531  | 1,1314    | 6,3101      | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 37 | -2,30259 | -2,30259  | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 40 | *        | 3,49953   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 41 | 0,74194  | 0,09531   | *           | *        | 0       | 0        | 0     | 0       | 1      | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 42 | 3,04927  | 3,40453   | *           | 1,46649  | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 44 | 2,89591  | 4,17592   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 45 | -2,30259 | -2,30259  | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 47 | *        | *         | *           | 2,78581  | 0       | 0        | 1     | 0       | 0      | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 48 | -2,30259 | 0,09531   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 1    | 0    | 0    | 0    | 1    | 2    |
| 49 | 0,09531  | 2,89591   | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 50 | 0,74194  | 6,21481   | 7,601       | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 51 | -2,30259 | -2,30259  | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 52 | *        | 1,1314    | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 53 | *        | *         | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 54 | 0,74194  | 0,74194   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 55 | 2,40695  | 3,69138   | *           | *        | 0       | 0        | 1     | 0       | 0      | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 56 | *        | *         | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 57 | -2,30259 | 5,20456   | 8,0064      | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 58 | -2,30259 | 3,40453   | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 59 | 0,74194  | 1,1314    | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 60 | *        | *         | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 61 | *        | *         | *           | 0,79615  | 0       | 0        | 1     | 0       | 0      | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 62 | -2,30259 | 0,09531   | *           | *        | 0       | 0        | 1     | 0       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 63 | -2,30259 | -2,30259  | -2,3026     | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |
| 64 | 1,41099  | *         | *           | *        | 0       | 0        | 0     | 1       | 0      | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    |

| AN | TACCEX | TACCFI | TACCGAC | TACCREL | TACTDO/C | TACTPRO | TACTSTO | TACTTRANS | CGEXT | CGHUM | CGIMP | CGMEC | CGVREA | CGINST | CGPROC |
|----|--------|--------|---------|---------|----------|---------|---------|-----------|-------|-------|-------|-------|--------|--------|--------|
| 1  | 0      | 0      | 1       | 1       | 1        | 0       | 0       | 0         | 0     | 1     | 0     | 0     | 1      | 0      | 0      |
| 3  | 1      | 0      | 0       | 0       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 0     | 1      | 0      | 0      |
| 6  | 1      | 0      | 0       | 0       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 8  | 0      | 0      | 0       | 1       | 1        | 0       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 10 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 12 | 0      | 1      | 0       | 0       | 1        | 0       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 13 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 14 | 1      | 0      | 1       | 0       | 0        | 0       | 0       | 0         | 0     | 0     | 0     | 0     | 0      | 0      | 0      |
| 15 | 0      | 0      | 1       | 1       | 1        | 0       | 0       | 0         | 0     | 0     | 1     | 0     | 0      | 0      | 0      |
| 16 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 1     | 0     | 1     | 0      | 0      | 0      |
| 17 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 19 | 0      | 0      | 0       | 1       | 0        | 1       | 0       | 0         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 20 | *      | *      | *       | *       | 0        | 0       | 1       | 0         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 21 | 0      | 1      | 0       | 1       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 23 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 25 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 26 | 1      | 0      | 0       | 1       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 27 | *      | *      | *       | *       | 0        | 0       | 1       | 0         | 0     | 0     | 0     | 0     | 1      | 0      | 0      |
| 32 | 0      | 0      | 0       | 1       | 1        | 0       | 0       | 0         | 0     | 1     | 0     | 0     | 0      | 0      | 0      |
| 33 | 0      | 0      | 1       | 1       | 1        | 0       | 0       | 0         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 37 | 0      | 0      | 1       | 1       | 0        | 1       | 0       | 0         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 40 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 41 | 1      | 0      | 0       | 0       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 42 | 1      | 0      | 0       | 1       | 0        | 1       | 0       | 0         | 1     | 0     | 0     | 0     | 0      | 0      | 0      |
| 44 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 45 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 1     | 0     | 0     | 0      | 0      | 0      |
| 47 | 1      | 1      | 0       | 0       | 0        | 1       | 0       | 0         | 0     | 1     | 0     | 1     | 0      | 0      | 0      |
| 48 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 1     | 0     | 0      | 0      | 0      |
| 49 | 0      | 0      | 0       | 1       | 1        | 0       | 0       | 0         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 50 | 0      | 0      | 1       | 1       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 51 | 0      | 0      | 0       | 1       | 0        | 1       | 0       | 0         | 0     | 1     | 0     | 1     | 0      | 0      | 0      |
| 52 | 0      | 0      | 0       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 53 | 0      | 0      | 0       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 54 | 0      | 0      | 0       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 55 | 1      | 0      | 0       | 0       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 56 | 0      | 1      | 0       | 1       | 1        | 0       | 0       | 0         | 0     | 0     | 1     | 0     | 0      | 0      | 0      |
| 57 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 58 | 0      | 0      | 1       | 1       | 0        | 0       | 0       | 1         | 0     | 1     | 0     | 0     | 0      | 1      | 0      |
| 59 | 1      | 0      | 0       | 0       | 0        | 1       | 0       | 0         | 0     | 0     | 0     | 1     | 0      | 1      | 0      |
| 60 | 1      | 1      | 0       | 0       | 0        | 1       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 61 | 1      | 1      | 0       | 0       | 0        | 0       | 0       | 0         | *     | *     | *     | *     | *      | *      | *      |
| 62 | 0      | 0      | 0       | 1       | 0        | 0       | 0       | 1         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 63 | 0      | 0      | 0       | 1       | 0        | 0       | 1       | 0         | 0     | 0     | 0     | 1     | 0      | 0      | 0      |
| 64 | 0      | 0      | 0       | 1       | *        | *       | *       | *         | 0     | 0     | 1     | 0     | 0      | 0      | 0      |



| AN | CGSERV | InQsus | KILLED | INJURED | EVACUATED | Damage | QSUS | Mdesenv | Menviadesenv | DÈCADA | ANTIGUITAT | ns*efs | ns*efl | ns*efg |
|----|--------|--------|--------|---------|-----------|--------|------|---------|--------------|--------|------------|--------|--------|--------|
| 1  | 0      | *      | 0      | 5       | *         | *      | *    | 1       | 0            | 8      | 1          | 0      | 0      | 2      |
| 3  | 0      | *      | *      | *       | *         | *      | *    | 1       | 0            | 8      | 1          | 0      | 0      | 0      |
| 6  | *      | *      | 7      | *       | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 8  | *      | *      | 5      | 85      | *         | *      | *    | 0       | 0            | 7      | 1          | 0      | 0      | 1      |
| 10 | 0      | 5,8582 | *      | *       | 10000     | *      | 350  | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 12 | *      | *      | *      | *       | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 13 | 0      | *      | *      | 17      | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 14 | 0      | *      | 0      | 10      | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 15 | 0      | *      | 0      | 24      | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 16 | 0      | *      | 3      | 20      | *         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 1      |
| 17 | 0      | *      | 0      | 0       | *         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 19 | 0      | 0,7419 | *      | 1       | *         | *      | 2    | 1       | 0            | 7      | 1          | 0      | 0      | 1      |
| 20 | 0      | *      | 0      | 0       | *         | *      | *    | 0       | 0            | 7      | 1          | 0      | 0      | 1      |
| 21 | *      | *      | *      | 0       | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 1      |
| 23 | 0      | *      | *      | *       | *         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 25 | 0      | 2,7147 | *      | *       | *         | *      | 15   | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 26 | *      | *      | *      | *       | *         | *      | *    | 0       | 0            | 5      | 0          | 0      | 0      | 1      |
| 27 | 0      | 6,3649 | *      | 4       | *         | *      | 581  | 1       | 0            | 5      | 0          | 0      | 0      | 1      |
| 32 | 0      | *      | 1      | *       | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 33 | 0      | *      | 1      | 3       | 550       | *      | *    | 1       | 0            | 8      | 1          | 0      | 0      | 1      |
| 37 | 0      | *      | 0      | 0       | *         | *      | *    | 1       | 0            | 8      | 1          | 0      | 0      | 1      |
| 40 | 0      | 3,2229 | *      | 33      | *         | *      | 25   | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 41 | *      | *      | 2      | 1       | *         | *      | *    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 42 | 0      | *      | 21     | 30      | *         | 4,234  | *    | 0       | 0            | 7      | 1          | 0      | 0      | 1      |
| 44 | 0      | 3,6402 | 18     | 65      | *         | *      | 38   | 0       | 1            | 7      | 1          | 0      | 0      | 0      |
| 45 | 0      | 1,411  | 0      | 0       | *         | *      | 4    | 1       | 0            | 7      | 1          | 0      | 0      | 0      |
| 47 | 0      | *      | *      | *       | *         | 16,113 | *    | 1       | 0            | 7      | 1          | 0      | 0      | 1      |
| 48 | 0      | 3,8733 | 0      | 1       | *         | *      | 48   | 1       | 0            | 8      | 1          | 0      | 0      | 0      |
| 49 | 0      | 1,1314 | 1      | 18      | *         | *      | 3    | 1       | 0            | 7      | 1          | 0      | 0      | 1      |
| 50 | *      | *      | 2      | 500     | 2000      | *      | *    | 0       | 1            | 7      | 1          | 0      | 0      | 1      |
| 51 | 0      | 2,3125 | 0      | 0       | *         | *      | 10   | 1       | 0            | 8      | 1          | 0      | 0      | 1      |
| 52 | 0      | *      | *      | 3       | *         | *      | *    | 1       | 0            | 5      | 0          | 0      | 0      | 0      |
| 53 | 0      | 3,0007 | *      | *       | *         | *      | 20   | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 54 | 0      | *      | 2      | 2       | *         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 55 | *      | *      | 11     | 40      | *         | *      | *    | 1       | 0            | 5      | 0          | 0      | 0      | 1      |
| 56 | 0      | *      | *      | *       | *         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 1      |
| 57 | 0      | 3,4045 | 0      | 182     | 3000      | *      | 30   | 0       | 1            | 8      | 1          | 0      | 0      | 0      |
| 58 | 0      | 3,8089 | 0      | 30      | *         | *      | 45   | 1       | 0            | 8      | 1          | 0      | 0      | 0      |
| 59 | 0      | *      | 2      | 3       | *         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 60 | *      | *      | *      | *       | *         | *      | *    | 0       | 0            | 8      | 1          | 0      | 0      | 1      |
| 81 | *      | *      | *      | *       | *         | 2,117  | *    | 1       | 0            | 7      | 1          | 0      | 0      | 1      |
| 82 | 0      | -1,204 | 0      | 1       | *         | *      | 0    | 1       | 0            | 7      | 1          | 0      | 0      | 1      |
| 63 | 0      | *      | 0      | 0       | 0         | *      | *    | 1       | 0            | 6      | 0          | 0      | 0      | 0      |
| 64 | 0      | 1,1314 | 4      | *       | *         | *      | 3    | 1       | 0            | 6      | 0          | 0      | 0      | 0      |



| AN | ns*tacre | ns*tatd/c | ns*tatpro | ns*tatsto | ns*tatran | ns*cgext | ns*cghu | ns*cgimp | ns*cgmec | ns*cgvre | ns*cginst | ns*cgproc | ns*cgserv |
|----|----------|-----------|-----------|-----------|-----------|----------|---------|----------|----------|----------|-----------|-----------|-----------|
| 1  | 2        | 2         | 0         | 0         | 0         | 0        | 2       | 0        | 0        | 2        | 0         | 0         | 0         |
| 3  | 0        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 0        | 1        | 0         | 0         | 0         |
| 6  | 0        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 8  | 1        | 1         | 0         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 10 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 12 | 0        | 1         | 0         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 13 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 14 | 0        | 0         | 0         | 0         | 0         | 0        | 0       | 0        | 0        | 0        | 0         | 0         | 0         |
| 15 | 1        | 1         | 0         | 0         | 0         | 0        | 0       | 1        | 0        | 0        | 0         | 0         | 0         |
| 16 | 1        | 0         | 0         | 0         | 1         | 0        | 1       | 0        | 1        | 0        | 0         | 0         | 0         |
| 17 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 19 | 1        | 0         | 1         | 0         | 0         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 20 | *        | 0         | 0         | 1         | 0         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 21 | 1        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 23 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 25 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 26 | 1        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 27 | *        | 0         | 0         | 1         | 0         | 0        | 0       | 0        | 0        | 1        | 0         | 0         | 0         |
| 32 | 1        | 1         | 0         | 0         | 0         | 0        | 1       | 0        | 0        | 0        | 0         | 0         | 0         |
| 33 | 1        | 1         | 0         | 0         | 0         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 37 | 1        | 0         | 1         | 0         | 0         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 40 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 41 | 0        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 42 | 1        | 0         | 1         | 0         | 0         | 1        | 0       | 0        | 0        | 0        | 0         | 0         | 0         |
| 44 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 45 | 1        | 0         | 0         | 0         | 1         | 0        | 1       | 0        | 0        | 0        | 0         | 0         | 0         |
| 47 | 0        | 0         | 1         | 0         | 0         | 0        | 1       | 0        | 1        | 0        | 0         | 0         | 0         |
| 48 | 2        | 0         | 0         | 0         | 2         | 0        | 0       | 2        | 0        | 0        | 0         | 0         | 0         |
| 49 | 1        | 1         | 0         | 0         | 0         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 50 | 1        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 51 | 1        | 0         | 1         | 0         | 0         | 0        | 1       | 0        | 1        | 0        | 0         | 0         | 0         |
| 52 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 53 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 54 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 55 | 0        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 56 | 1        | 1         | 0         | 0         | 0         | 0        | 0       | 1        | 0        | 0        | 0         | 0         | 0         |
| 57 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 58 | 1        | 0         | 0         | 0         | 1         | 0        | 1       | 0        | 0        | 0        | 1         | 0         | 0         |
| 59 | 0        | 0         | 1         | 0         | 0         | 0        | 0       | 0        | 1        | 0        | 1         | 0         | 0         |
| 60 | 0        | 0         | 1         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 61 | 0        | 0         | 0         | 0         | 0         | *        | *       | *        | *        | *        | *         | *         | *         |
| 62 | 1        | 0         | 0         | 0         | 1         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 63 | 1        | 0         | 0         | 1         | 0         | 0        | 0       | 0        | 1        | 0        | 0         | 0         | 0         |
| 64 | 1        | *         | *         | *         | *         | 0        | 0       | 1        | 0        | 0        | 0         | 0         | 0         |

| AN | lqs*efs | lqs*efl | lqs*efg | lqs*efp | lqs*efd |
|----|---------|---------|---------|---------|---------|
| 1  | *       | *       | *       | *       | *       |
| 3  | *       | *       | *       | *       | *       |
| 6  | *       | *       | *       | *       | *       |
| 8  | *       | *       | *       | *       | *       |
| 10 | 0       | 0       | 0       | 5,8582  | 0       |
| 12 | *       | *       | *       | *       | *       |
| 13 | *       | *       | *       | *       | *       |
| 14 | *       | *       | *       | *       | *       |
| 15 | *       | *       | *       | *       | *       |
| 16 | *       | *       | *       | *       | *       |
| 17 | *       | *       | *       | *       | *       |
| 19 | 0       | 0       | 0,7419  | 0       | 0       |
| 20 | *       | *       | *       | *       | *       |
| 21 | *       | *       | *       | *       | *       |
| 23 | *       | *       | *       | *       | *       |
| 25 | 0       | 0       | 0       | 2,7147  | 0       |
| 26 | *       | *       | *       | *       | *       |
| 27 | 0       | 0       | 6,3649  | 0       | 0       |
| 32 | *       | *       | *       | *       | *       |
| 33 | *       | *       | *       | *       | *       |
| 37 | *       | *       | *       | *       | *       |
| 40 | 0       | 0       | 0       | 3,2229  | 0       |
| 41 | *       | *       | *       | *       | *       |
| 42 | *       | *       | *       | *       | *       |
| 44 | 0       | 0       | 0       | 3,6402  | 0       |
| 45 | 0       | 0       | 0       | 1,411   | 0       |
| 47 | *       | *       | *       | *       | *       |
| 48 | 0       | 0       | 0       | 3,8733  | 0       |
| 49 | 0       | 0       | 1,1314  | 0       | 0       |
| 50 | *       | *       | *       | *       | *       |
| 51 | 0       | 0       | 2,3125  | 0       | 0       |
| 52 | *       | *       | *       | *       | *       |
| 53 | 0       | 0       | 0       | 3,0007  | 0       |
| 54 | *       | *       | *       | *       | *       |
| 55 | *       | *       | *       | *       | *       |
| 56 | *       | *       | *       | *       | *       |
| 57 | 0       | 0       | 0       | 3,4045  | 0       |
| 58 | 0       | 0       | 0       | 3,8089  | 0       |
| 59 | *       | *       | *       | *       | *       |
| 60 | *       | *       | *       | *       | *       |
| 61 | *       | *       | *       | *       | *       |
| 62 | 0       | 0       | -1,204  | 0       | 0       |
| 63 | *       | *       | *       | *       | *       |
| 64 | 0       | 0       | 0       | 1,1314  | 0       |

## BIBLIOGRAFIA

BIRKES, D. i DODGE, Y. (1.993). *Alternative Methods of Regression*, Wiley.

DRAPER, N.R. i SMITH, H. (1.998). *Applied Regression Analysis*, 3<sup>rd</sup> Edition, Wiley.

McCULLAGH, P. i NELDER, J. (1.987). *Generalized linear models*, 2<sup>nd</sup> Edition, Chapman & Hall.

*MHIDAS. Winspirs user's manual* (1.996), Silverplatter International, London.

PEÑA, D. (1.987). Estadística, *Modelos y Métodos 2. Modelos Lineales y Series Temporales*, Alianza Universidad Textos.

ROUSEWOU, P. J. i LEROY, A. M. (1.986). *Robust Regression and Outlier Detection*, Wiley.

SÁNCHEZ NAVARRO, D. i CARBONELL AYUSO, I. (1.997). *Microsoft Access. Iniciación y Referéncia*, Osborne/McGraw-Hill.

SOLER CHICA, Enrique (1.997). *Microsoft Excel. Iniciación y Referéncia*, Osborne/McGraw-Hill.

WEISBERG, S. (1.985). *Applied Linear Regression*, 2<sup>nd</sup> Edition, Wiley.